Connection Admission Control and Packet Scheduling for IEEE 802.16 Networks

Prepared by:

Samuel Kehinde Falowo

Supervised by:

Neco Ventura

This thesis is submitted in partial fulfillment of the academic requirements
for the degree of
Master of Science in Electrical Engineering
in the Faculty of Engineering and The Built Environment
University of Cape Town
May 2012

As the candidate's supervisor, I have approved this dissertation for submission.

Name: Mr. Neco Ventura

Signed: _____

Date: _____

Declaration

I hereby declare that: (1) the above thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance of my supervisor, I have received no assistance apart from that stated below; (2) except as stated below, neither the substance or any part of the thesis has been submitted in the past, or is being, or is to be submitted for a degree in the University or any other University.

I am now presenting the thesis for examination for the Degree of MSc in Electrical Engineering. I also grant the University free license to reproduce the above thesis in whole or in part, for the purpose of research.

Samuel K. Falowo

Date

To the glory of God and the benefit of mankind.

Synopsis

The IEEE 802.16 standard introduced as one of the Wireless Metropolitan Area Networks (WMAN) for Broadband Wireless Access (BWA) which is known as Worldwide Interoperability for Microwave Access (WiMAX), provides a solution of broadband connectivity to areas where wired infrastructure is economically and technically infeasible. Apart from the advantage of having high speeds and low costs, IEEE 802.16 has the capability to simultaneously support various service types with required QoS characteristics. With increasing bandwidth demands from network users and emerging bandwidth-intensive applications, there is need for provisioning of mechanisms that will allow network users to access the network efficiently with adequate quality of service guarantees. While IEEE 802.16 standard defines medium access control (MAC) and physical (PHY) layers specification, admission control and packet scheduling mechanisms which are important elements of QoS provisioning are left to vendors to design and implement for service differentiation and QoS support.

This thesis focuses on admission control and packet scheduling for IEEE 802.16 networks and develops a connection admission control and packet scheduling mechanisms for service differentiation and QoS support in IEEE 802.16 networks. A Quadra-threshold bandwidth-based connection admission control is proposed as the main part of the thesis. In the connection admission control algorithm, each service type of a connection request is associated with a set threshold. The set threshold is used to differentiate and prioritize a service type according to its QoS requirement. To reduce connection blocking probability during congestion, a bandwidth degradation mechanism is developed to reduce the bandwidth of ongoing connections to their minimum requirements so that more connection requests can be admitted. The admission control mechanism ensures that connection requests of a service type are only admitted within the set threshold for that service type without overloading the network. A priority-based packet scheduling algorithm that adopts processor sharing mechanisms is developed to schedule connection for packet transmission. A mathematical analysis using 4-dimensional Markov decision process and queuing theory is used for algorithm development. The algorithm addresses the shortcomings in existing algorithms found in literature, such as complete partitioning of resources.

The performance of the proposed algorithm is evaluated using MATLAB. Five service

types as defined in IEEE 802.16 are considered. QoS performance is measured with connection-level QoS metrics namely connection blocking probability and throughput, as well as delay at the packet level. The performance is carried out under different loading conditions and the results obtained are analysed. Results analysis show that Quadra-threshold bandwidth-based connection admission control performs better than the generic scheme of complete partitioning and scheme without admission control in terms of blocking probability. In addition, more connection requests can be admitted into the network with QoS guarantee when bandwidth degradation is used in the proposed algorithm. The connection throughput shows that connection types with higher priority are given higher throughput for service differentiation with minimal delay. The proposed connection admission control and packet scheduling ensure QoS guarantee to network users.

Acknowledgements

I am grateful to the Lord almighty for His gift of life and daily blessings.

My sincere appreciation goes to my supervisor, Neco Ventura for his fatherly support and guidance during the period of my masters program.

I would like to thank my parents, sisters, brothers and my loved ones for their love and encouragement at all time, especially Dr. O.E Falowo for given me the opportunity to get this far.

I would like to thank my colleagues in the center for broadband networks and applications, especially Joyce Mwangama, Ernest Petro and Alien Ramboli for their friendly supports.

Abbreviations and acronyms

1G	first generation
2G	second generation
3G	third generation
4G	fourth generation
3GPP	third generation partnership project
AMPS	advanced mobile phone system
BBU	basic bandwidth unit
BE	best effort
BR	bandwidth request
BS	base station
BW	bandwidth
BWA	broadband wireless access
CDMA	code division multiple access
CID	connection identifier
CPS	common part sublayer
CS	convergence sublayer
DAMA	demand assigned multiple access
DL	downlink
DSL	digital subscriber line
EDGE	enhanced data for GSM evolution
ertPS	extended real time polling service
FDMA	frequency division multiple access

FTP	file transfer protocol
GPRS	general packet radio service
GSM	global system for mobile communications
HTTP	hyper text transfer protocol
IE	information element
IP	internet protocol
LAN	local area network
LOS	line-of-sight
MAC	medium access control
MAN	metropolitan area network
MCS	modulation coding scheme
MS	mobile station
NLOS	non-line-of-sight
nrtPS	non-real-time polling service
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
PAN	personal area network
PDU	protocol data unit
PHY	physical layer
PMP	point-to-multipoint
PS	physical slot
QoS	quality of service
RTG	receive/transmit transition gap
rtPS	real-time polling service

SC	single carrier
SDU	service data unit
SF	service flow
SFID	service flow identifier
SOHO	small office/home office
SS	subscriber station
SSTG	subscriber station transition gap
TDD	time division duplex or duplexing
TDM	time division multiplexing
TDMA	time division multiple access
TTG	transmit/receive transition gap
UGS	unsolicited grant service
UL	uplink
UMTS	universal mobile telecommunications system
VoIP	voice over IP
W-CDMA	wideband CDMA
WirelessMAN	Wireless Metropolitan Area Networks
WirelessHUMAN	Wireless High-speed Unlicensed Metropolitan Area Networks
Wi-Fi	wireless fidelity
WiMAX	worldwide interoperability for microwave access
WLAN	wireless local area network

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Emerging telecommunications services and applications are the strong drivers of increasing bandwidth demands for last mile broadband access. They pose new requirements to the existing network access technologies [1]. The demand for Internet Protocol (IP) connectivity is yielding rapid development in the wireless access network domain due to the proliferation of portable multimedia application devices such as Laptop Computers, Smartphone, Hand-held Computers and Tablet Personal Computers. While the deployment of these wireless devices tends to address issues like network accessibility and portability they also have their challenges. Although subscribers are willing to pay more for more bandwidth to obtain better quality of service (QoS) from network operators, limited bandwidth and allocation of the available bandwidth among different subscribers are big challenges.

Users' expectations are continuously increasing with regard to the variety of services and applications across a range of devices. There is a need to support these services and applications by available radio access technologies efficiently in order to guarantee good quality of service. Different radio access technologies have been developed and deployed for efficient network usage, installation and operation and new ones are still in development for standardization. Even though the available radio access technologies such as Global System for Mobile Communications (GSM), Wideband Code Division Multiple Access (W-CDMA), Bluetooth and Wireless Fidelity (Wi-Fi) have been widely deployed and used, they are not without their limitations. They have low data rates, limited coverage and inability to support different service types simultaneously.

The insufficient throughput support for broadband IP traffic in the existing wireless radio access technologies motivated the 3rd Generation Partnership Project (3GPP) to account WiMAX as a complementary broadband wireless access (BWA) technology. Although cable and digital subscriber line (DSL) are already deployed on a large scale, IEEE 802.16, also known as Worldwide Interoperability for Microwave Access (WiMAX) [2], [3], [4] is emerging as an access technology with several advantages. These include wireless connectivity, flexible base station architecture and a number of subscriber stations under the antenna sector of the base station. A base station is responsible for performing access control and radio resource allocation to the subscriber stations. WiMAX also offers a wide area of coverage up to tens of kilometres in

line of sight (LOS) environment, support for non LOS operation, high capacity and data rates of up to 70Mbps. In addition, WiMAX provides a high level of security with support for advanced encryption standard (AES) and triple data encryption standard (3DES). It also provides QoS support for real time data streams, mobility support, easy and inexpensive deployment and flexibility in spectrum allocation in licensed and unlicensed frequency bands [5].

The cost of backhaul for cellular and wireless fidelity (Wi-Fi) networks represents the substantial part of their recurrent cost. WiMAX can support different types of services and operators can use the WiMAX equipment to provide hotspots and backhaul for their networks and high-speed enterprise connectivity for business customers. Furthermore, WiMAX can also provide video surveillance cameras with broadband connectivity to control centres in real time. Two of the important components of IEEE 802.16 architecture that handle how radio resources are distributed among users and ensure that users are given their negotiated QoS are not defined in the 802.16 standard but left to vendors to design and implement [6]. These components, connection admission control and packet scheduling are vital mechanisms of QoS provisioning in WiMAX networks.

Connection Admission Control (CAC) scheme provides users with access to a wireless network with the objectives of providing services to users with guaranteed QoS by limiting the number of users accessing the network and at the same time achieving efficient resource utilization. Scheduling schemes are used to resolve contention for shared resources in a network by allocating bandwidth to users and determining their transmission priority. A scheduler allocates resources and establishes the order in which information flows are served ensuring that the QoS requirements for each information packet are guaranteed. The allocated resources include bandwidth; that determines the rate at which packet is transmitted, priority; that determines which packet is transmitted first and packet buffering; which determines the memory space reserved for storing packets awaiting transmission.[7].

Connection admission Control and Packet Scheduling are important elements of Quality of Service provisioning for IEEE 802.16 networks. When these components are efficiently designed and implemented in the IEEE 802.16 architecture, there will be an improvement in system performance and QoS as perceived by network users.

1.1 Research Motivation

Network users expect to be able to connect to anyone, anywhere at any time using any device with QoS guaranteed. The cellular technology that is targeted at providing voice services has limited data rate that is around 10 Mbps or lower, and does not scale to the capacity of all-IP media-centric network [8]. The success of IP services deployment requires true mobile broadband IP connectivity on a global scale. WiMAX is a good candidate for all-IP technology with the aim of providing voice, data, video and multimedia services at high speeds while remaining cost effective. With the increase in the emergence of users' application devices with different QoS requirements to compete for limited network resource, implementation of QoS guarantee in such a network is a requirement for efficient support of applications over such networks.

Some radio access technologies, such as WCDMA-HSPA, CDMA 2000-EVDO [9], 802.15 PAN (Personal Area Network) [10] and 802.11 WLAN (Wireless Local Area Network) [11] have emerged to meet the challenge of growing demand for high bandwidth application both in fixed and mobile environment. WiMAX does not only provide high bandwidth but also combines the advantages of WLAN and cellular networks to meet the challenge of integrating different types of services such as voice, video, data and internet. The current state of IEEE 802.16 standards for local and metropolitan area networks, IEEE 802.16-2009 [4] merges the former IEEE 802.16d air interface specifications for fixed broadband wireless access systems and IEEE 802.16e air interface specifications for mobile broadband wireless access systems. The long-term evolution of WiMAX will achieve 100Mbits/s mobile and 1Gbit/s fixed-nomadic bandwidth rate that will be a viable alternative to 4G next generation mobile network technologies [12].

Furthermore, WiMAX technology is new and still under development [13]. IEEE 802.16 Standard defines the PHY and MAC specifications for WiMAX, however, connection admission control (CAC) and packet scheduling (PS) are left to the manufactures or vendors to implement [6], [14] and [15]. Research in this area has only started to gain ground.

In addition, existing works of CAC and packet scheduling on WiMAX have some drawbacks. The designed algorithms are complex, which may be infeasible to implement. Simple connection admission control and packet scheduling algorithms will ensure efficient resource allocation among different service types. Much of the related literature did not incorporate all the

service types defined in 802.16 standards. IEEE 802.16 defined five service types with different QoS requirements. Each service type needs to be treated with peculiarity to guarantee the negotiated QoS.

Moreover, ideal connection admission control and packet scheduling algorithms are necessary for QoS guarantee, because, they have significant effect on the performance of the system. When admission of users into a network is not controlled, the network becomes overloaded and system performance drops. The reduction in performance of the system will not only affect the newly admitted users but also ongoing users will be affected. The users' perspectives of system performance is very important. These are determined by the quality of service received by the users. When admission control and packet scheduling are used, the number of users present in the system at any given time will be restricted to the capacity that the system can handle efficiently, hence providing better quality of service.

1.1.1 Problem Definition

The growing need for wireless broadband access for different emerging user applications has increased the need for QoS guarantee and radio resource utilization. Recent studies have shown that the proportion of VoIP users continued to grow from 28% of users in 2008 (up from 20% of user in 2007) to more than 50% in 2010 [16]. The continuous growth in users' application devices coupled with limited radio resources available for users has increased the need for high QoS and better resource utilization. Without efficient CAC and Packet Scheduling, networks will not be able to provide QoS guarantee to real time applications like voice and video and efficiently utilize the network resources [17]. It is observed that existing wire line and wireless schedulers do not perform very well with respect to different scheduling classes defined in the WiMAX Standard. In addition, each of this traffic classes has a different scheduling requirement and consequently, it has become necessary to design appropriate scheduling frameworks [18]. The problem of ensuring QoS is basically that of how to allocate available resources among users in order to meet QoS requirements [19].

The significance of CAC and Packet Scheduling is to ensure that when a user application is given network resources, the QoS requirements of such application are guaranteed. The contribution of this work is to develop a standards compliant CAC and Packet Scheduling

algorithm that will meet the QoS requirements of scheduling services. A simple and efficient algorithm would be developed for the five scheduling services namely: Unsolicited Grant Service (UGS), real time polling service (rtPS), extended real time polling service (ertPS), non-real time polling service (nrtPS) and best effort (BE) service would be considered.

1.1.2 Research Questions

The research questions for this thesis can be summarised as follows;

1. In what ways can the admission probability of high priority connections be increased? Different service types with different quality of service requirements are defined in the IEEE 802.16 standards. The high priority connections need to be given sufficient share of the network resources so that their quality of service can be guaranteed.
2. What is the best way to improve fairness and throughput in resource allocation to different services types? Fairness measures or metrics are used to determine whether users or applications are receiving a fair share of network resources. If resource allocation is fair to service flows and flows are allocated their required bandwidth, the required data rate delivery i.e. throughput will be guaranteed.
3. In what ways can the delay requirement of real time applications be improved? Packet delay in VoIP application is not desirable since it renders such application meaningless. Satisfying delay requirement will improve the QoS rendered to network users.
4. By what means can the bandwidth assigned to uplink subframe be efficiently utilized? If the uplink bandwidth is efficiently utilized more users will be able to access the networks.

1.2 Thesis Objectives

Quality of service guarantee for different service types in WiMAX networks will depend on efficient management of radio resources such as bandwidth. This thesis proposes a connection admission control and packet scheduling schemes to manage radio resources in IEEE 802.16 networks. Algorithms for connection admission control and packet scheduling will be designed to efficiently manage network resources among the different service types designed in the IEEE 802.16 Standards.

The objectives of this thesis can be summarized as follows:

- Carry out comprehensive review of existing literatures on admission control and packet scheduling for wireless networks. Identify their benefits and limitations and how these limitations can be improved upon. Analysis of related work done will give better insight on suitable approach to address the thesis.
- Examine the key requirements of CAC and packet scheduling for efficient resource utilization and quality of service guarantee. The IEEE 802.16 architecture will be examined to identify the basic components that should be introduced, modified or improved to incorporate the proposed schemes.
- Design algorithms to perform CAC and packet scheduling in 802.16 networks and ensure that QoS of different service types are considered. The design algorithm of a system is an important factor on performance of the system.
- Identify the parameters of interest for evaluation and how these parameters affect system performance. It is important to examine parameters of interest at connection level and packet level in terms of connection blocking probability, throughput and packet drop rate.
- Analyse the performance of the proposed schemes under different network loads.

1.3 Scope and Limitations

The IEEE 802.16 standard was introduced in 2001 to address line of sight (LOS) access spectrum ranges from 10GHz to 63GHz frequency band. The Standard was extended in 2004 and formed the IEEE 802.16d-2004 [2], which defined the specification of the air interface for fixed broadband wireless system and is also referred to as "Fixed WiMAX".

The current state of the standard defines "Air Interface for Fixed and Mobile Broadband Wireless Access System" to support both fixed and mobile wireless communications and it is officially named IEEE 802.16-2009 [4]. In IEEE 802.16-2009, two operation modes are defined: Point-to-Multipoint (PMP) and Mesh Modes. A Point-to-Multipoint architecture consists of a base station (BS) and number of subscriber stations (SSs). The BS provides connectivity, management, control and centrally coordinates the SSs under its antenna sector. A base Station

is capable of handling multiple SSs and within a given frequency channel and antenna sector, all SSs receive the same transmission. In this thesis, point-to-multipoint transmission mode in 10-66 GHz bands is considered because, 10-66 GHz bands provide a physical environment where, due to short wavelength, line-of-sight (LOS) is required and multipath is negligible [4].

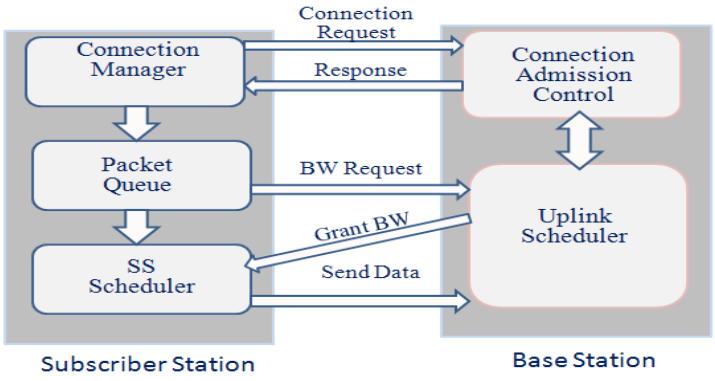

Figure 1.1. Architecture of IEEE 802.16 Networks

IEEE 802.16 architecture in Figure 1.1 shows request and response interactions between a base station and a subscriber station. The focus of this work is to design connection admission control and packet scheduling algorithms for connections types originating from subscriber stations to the base station. Connections from different service types are limited to the ones originating from subscriber stations within the antenna sector of base station. Therefore, mobility is within the sector and handoff connections are treated as new connection originating within the sector. Issues relating to signalling mechanisms and protocols are beyond the scope of this work.

Connection admission control and packet scheduling are defined in 802.16 architecture to be operated in the medium access control of protocol layer. Given that the multipath effect is negligible [3], we assume the perfect physical channel conditions to reduce the complexity of the algorithm. Therefore, the thesis work is focused on medium access control layer of the protocol stalk. Algorithm design is important for efficient connection admission control and packet scheduling which is the main aspect of this thesis.

1.4 Thesis Outline

The remainder of this thesis is organised as follows;

Chapter 2 gives a brief overview of IEEE 802.16 networks with PHY and MAC provisions defined in 802.16 standards. This chapter presents 802.16 QoS enhancements. The knowledge of 802.16 components and QoS enhancements is expected to form the basis of this thesis. In addition, a comprehensive review of literature works regarding connection admission control and packet scheduling are presented.

Chapter 3 presents the design considerations and requirements for the proposed connection admission control and packet scheduling for 802.16 networks. Design requirements for service types defined in 802.16 are considered.

Chapter 4 presents analytical frameworks and modelling used in the study. The architectural components of the proposed connection admission control and packet scheduling from MAC layer point of view are described. The metrics used to evaluate the performance of the proposed work are discussed.

Chapter 5 provides the performance test and results obtained. The results are thoroughly analysed and compared with the existing schemes.

Finally, chapter 6 presents a set of conclusions derive from the results obtained in the previous chapter. A summary of issues encountered in previous chapter are illustrated. In addition, this chapter gives some recommendations and future work.

1.5 Contributions

The major contributions of this research work are documented in the following peer reviewed conference publications:

[1] Samuel Falowo, Neco Ventura, "Connection Admission Control (CAC) for QoS Differentiation in PMP IEEE 802.16 Networks," Proceedings of the IEEE AFRICON Conference, The Fall and Resort Centre, Livingstone, Zambia, 13-15 September 2011, ISBN: 978-0-620-50893-3.

[2] Samuel Falowo, Neco Ventura, "An Efficient Connection Admission Control (CAC) for QoS Provisioning in IEEE 802.16," Proceedings of South Africa Telecommunication Networks and Applications Conference (SATNAC), East London Convention Centre, South

Africa, 4-7 September, 2011, ISBN: 978-1-61284-991-1.

Chapter 2 Background and Literature Review

Wireless technology is rapidly evolving and is playing an increasingly important role in the lives of people throughout the world. It is envisaged not only as the suitable technology to enable users to access the network but also viewed as the practical way to quickly construct the edge networks and even the core networks [20]. The proliferation of mobile devices has been very fast during the last years especially in the developing countries where for many people, the only access to internet is through the wireless telecom terminals [21]. Standards organisations have created numerous standards for wireless technologies. From the late 1970s until today, there have been different generations of wireless systems based on different access technologies namely; first generation (1G) wireless system, second generation (2G) wireless system, third generation (3G) wireless system [1] and the fourth generation (4G) wireless system is still under standardization.

Table 2.1: Basic Characteristics of Wireless Generation Systems

Features	**1G**	**2G**	**2.5G**	**3G**
Air Interfaces	FDMA	TDMA, CDMA	TDMA	W-CDMA, TD-CDMA, CDMA2000
Bandwidth		~ 10kbps	~ 100kbps	
Data Rate	No data	Circuit Switched	Packet Switched	Packet Switched
Example of Services	AMPS	GSM	GPRS, EDGE	UMTS, CDMA2000
Modulation	Analog	Digital	Digital	Digital
Voice Traffic	Circuit Switched	Circuit Switched	Circuit Switched	Packet Switched (VoIP)

Voice calls, internet browsing and sending of short messages that require low bit rates can be supported by the wireless technologies mentioned in Table 2.1. On the other hand, bandwidth intensive applications such as image, video and computer graphics require high bit rates which cannot be efficiently supported by these wireless technologies. Advanced wireless technologies are required to provide broadband wireless access that can simultaneously support

different applications with good quality of service. The recent wireless technologies that have been widely deployed are; Bluetooth referred to as Personal Area Network (PAN), and Wireless Fidelity (Wi-Fi) referred to as Wireless Local Area Network (WLAN). They both belong to IEEE 802 family of Standards.

Wireless PAN 802.15 Standards (Bluetooth)

Bluetooth network has no network infrastructure other than nodes [10]. Its distance ranges within 10 meters with provision of rapid ad hoc connections without cable and line of sight requirement. It operates at lower power levels than Wi-Fi with many devices transmitting at just 1 or 10 milli-watts. The total data rate is within the 1Mbps range. Being simpler in design, the entire set of Bluetooth can fit into a low-cost chip.

Wireless LAN 802.11 Standards (Wi-Fi)

The IEEE 802.11 WLAN network architecture consists of an access point (AP) that connects to other networks to provide access to different WLAN users [11]. Each AP in Wi-Fi has a finite range within which a wireless connection can be maintained between the client device and the AP. The IEEE 802.11 offers a shared maximum data rate of 11Mbps; the data rate can be extended to 54Mbps in 802.11g. In 802.11n, further increase in data rate to over 100Mbps is achieved. Wi-Fi was developed to be used for mobile computing devices, such as laptops in LANs, but it is now increasingly used to create a mesh network and connectivity in peer-to-peer ad hoc networks. WLAN provides flexible connectivity and robustness; however, it is limited in terms of coverage, QoS guarantee, safety and security and interference mitigation.

Fourth generation (4G) wireless systems offer a higher data rate of 50-200Mbps, converges with TV broadcast network, converges with fixed wireless system and is an end-to-end all IPv6 network. It will provide more wireless services which are seamless but are at low cost and provide tighter security. The suitable wireless technologies for 4G networks are Long Time Evolution (LTE) and IEEE 802.16 referred to as Worldwide Interoperability for Microwave Access (WiMAX). Third Generation Partnership Project (3GPP) is evolving towards OFDM and OFDMA in its LTE. Meanwhile, the IEEE 802.16-2009 [4] is using OFDM and OFDMA physical interfaces. Both interfaces support fixed and mobile wireless access in line of sight and non-line of sight and are simple and less expensive to deploy [20] .

2.1 IEEE 802.16 Wireless Metropolitan Area Networks (WMAN)

The IEEE 802.16 Forum describes WiMAX as a standard based technology enabling delivery of last mile wireless broadband access as an alternative to cable and digital subscriber line (DSL). WiMAX specifications provide symmetrical bandwidth up to 48 kilometres with support for different service types and quality of service guarantee. The IEEE 802.16 standard defines a flexible architecture of a base station (BS) and a number of subscriber stations (SSs) that operates in two modes: Point-to-Multipoint (PMP) and mesh modes. In PMP mode (see Figure 2.1), the downlink transmission is from a central BS to SSs. The BS provides connectivity, management, control and centrally coordinates the SS under its antenna sector. The

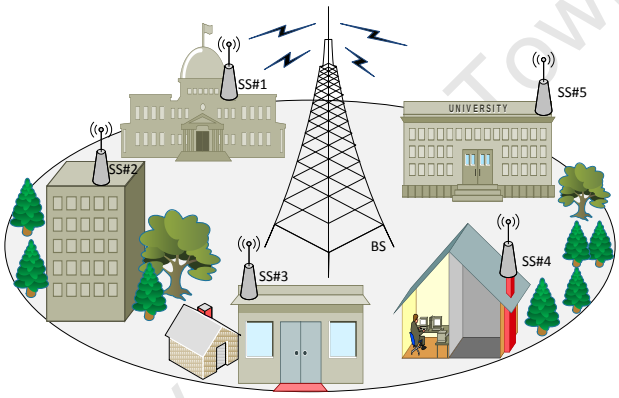

Figure 2.1: The PMP Operation Mode

BS is capable of handling multiple SSs and within a given frequency channel and antenna sector, all SSs receive the same transmission, or parts thereof. The link from SS to BS is called the uplink and the link from the BS to SS is called downlink. In the uplink transmission, the SSs share the uplink channel on a demand basis while the downlink channel is fully controlled by BS. Transmission opportunities are issued to the SSs by the BS based on transmission requests from the user. In mesh mode of Figure 2.2, SSs may have no direct link to BS. The transmission between SSs and BS can be established in two ways; centralized and distributed.

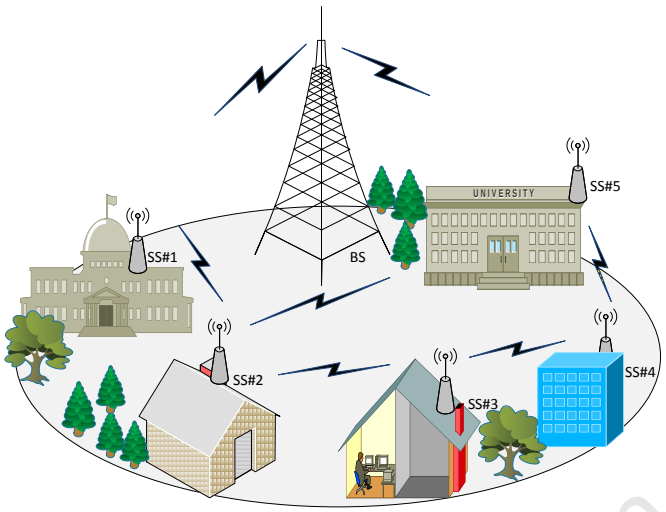

Figure 2.2: The Mesh Operation Mode

In centralized operation, BS manages the control of the network and the uplink and downlink bandwidths. In a distributed manner, all SSs referred to as nodes periodically exchange their schedules and bandwidth requests/grants and then come up with a suitable communication schedule running the distributed algorithm installed in every node.

The main difference between the PMP and mesh modes is that in the PMP mode, traffic only occurs between the BS and SSs; while in mesh mode traffic can be routed through other SSs and can occur directly between SSs.

To allow flexible spectrum usage, Time Division Duplex (TDD) and Frequency Division Duplex (FDD) are supported in WiMAX. IEEE 802.16 Standards define channel access modes in uplink and downlink directions. Time Division Multiplexing (TDM), Time Division Multiple Access (TDMA), Demand Assigned Multiple Access (DAMA), Orthogonal Frequency Division Multiplexing (OFDM) and Orthogonal Frequency Division Multiple Access (OFDMA) are the channel access modes defined in the Standards. The Standards also specify the medium access control layer (MAC) and physical layer (PHY) of fixed and mobile point-to-multipoint (PMP) broadband wireless access (BWA) system with various services. The MAC is designed to support various physical layer (PHY) specifications with each PHY specification structured to suit a particular operational environment.

The IEEE 802.16 [4] defines three PHY specifications to address line of sight (LOS) and non-LOS (NLOS) spectrum access in the licensed and unlicensed frequency bands. The first

standard was developed in 2001 to address LOS access spectrum ranges from 10GHz to 63GHz frequency band. The standard was extended in 2004 to form the IEEE 802.16d-2004 air interface specification for fixed broadband wireless access system, which is also known as "Fixed WiMAX" [2]. Mobility support was introduced into the standard in 2005 named IEEE 802.16e, the "Mobile WiMAX" [22] . The actual state of the standard defines "Air Interface for Fixed and Mobile Broadband Wireless Access System" to support both fixed and mobile wireless communications and it is officially named IEEE 802.16-2009 [4].

The IEEE 802.16-2009 PHY defines three specifications that are suited for a different operational environment according to the radio frequency band in which each specification operates. The radio PHY specifications are:

- 10-66 GHz licensed band
- Frequencies below 11 GHz and
- Licensed-exempt frequencies below 11GHz (primarily 5-6 GHz).

The 10-66 GHz licensed band which is also known as the "WirelessMAN-SC" air interface where SC means single carrier modulation is defined for point-to-multipoint (PMP) channel access where due to short wavelength, line of sight is required and multipath is negligible. The typical channel bandwidth is 25 MHz or 28MHz and the raw data rates can exceed 120 Mb/s. Point-to-multipoint channel access can be used to serve applications like small office/home office (SOHO) or larger office areas. Two specifications were defined for the frequencies below 11 GHz: "WirelessMAN-OFDM" and "WirelessMAN-OFDMA" with near-LOS and NLOS supports. In this physical environment LOS is not required and multipath may be significant. The common approaches to mitigate the inter-symbol interference (ISI) caused by multipath propagation are orthogonal frequency-division multiplexing (OFDM) and orthogonal frequency-division multiple access (OFDMA). The specification for license-exempt frequency bands below 11 GHz is similar to that of licensed frequency bands below 11 GHZ except that the introduction of co-existence issue and additional interference, which must be prevented from other users. The IEEE 802.16 standard defines MAC and PHY layers requirements. While connection admission control and packet scheduling mechanisms are not stated in 802.16 standards, these components can be implemented in the medium access control (MAC) layer; therefore, it is very important to look at some features of MAC layer for better understanding.

2.1.1 Medium Access Control (MAC) Layer

The physical medium is not informed of quality of service (QoS) requirements and is not aware of the nature of the application, such as VoIP, HTTP or FTP. The MAC, which resides above the PHY, is responsible for controlling and multiplexing various links over the same physical medium.

The MAC layer of WiMAX is divided into three distinct components (see Figure 2.3);

A. The Service-specific Convergence Sublayer (CS)

B. The Common Part Sublayer (CPS)

C. The Security Sublayer

A. The Service-specific Convergence Sublayer (CS)

Medium access control CS can be viewed as an adaptation layer that separates the higher-level network protocols from the rest of the WiMAX MAC and physical layers (see Figure 2.3). Convergence sublayer receives higher layer (Protocol Data Units) PDUs, classifies and associates them to a proper MAC service flow. The higher layer PDU is also known as MAC Service Data Unit (SDU). Classification is the process by which MAC SDU is mapped onto a particular service flow where a set of parameters are defined to ensure QoS and transmission between the MAC peers of the BS and the SS. The convergence sublayer is also responsible for the operations such as packet header compression and reconstruction.

A. The Common Part Sublayer (CPS)

MAC common part sublayer resides in the middle of the MAC layer (see Figure 2.3). The classified SDUs arrive at the MAC CPS where they are assembled into PDU, which is the basic payload unit of the WiMAX network. Several SDUs may be packed into a single SDU or a single SDU may be fragmented into several PDUs. The MAC CPS at the receiving end does the opposite operation to extract the SDUs, which are delivered to the higher layers. The MAC CPS provides the core functionality of the system access, bandwidth allocation, connection establishment, and connection maintenance.

A. The Security Sublayer

The security Sublayer is responsible for encryption, authorization, and proper exchange of encryption keys between the BS and the SS.

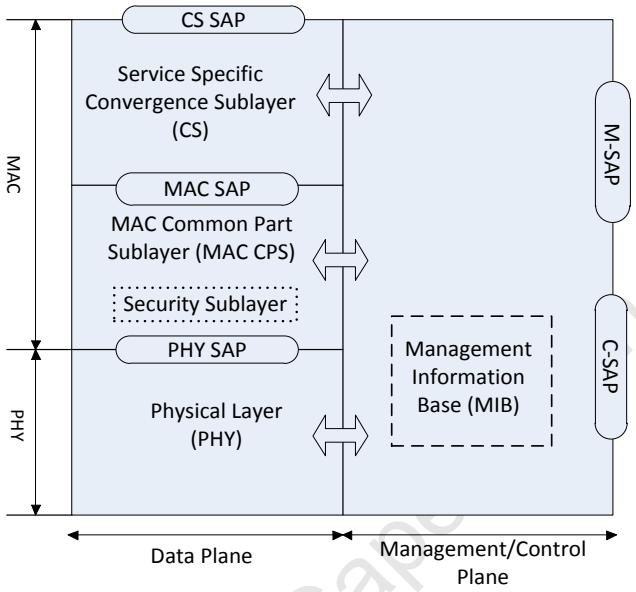

Figure 2.3: IEEE Standard 802.16 Protocol Reference Model

2.1.2 Physical (PHY) Layer

In a network, the purpose of the PHY layer is to reliably deliver information bits from the transmitter to the receiver, using the physical medium. The WirelessMAN-SC PHY specification supports operation in the 10-66GHz frequency band. In order to allow for flexible spectrum usage, both TDD and FDD transmission modes are supported. Both modes support burst profiling in which transmission parameters, including modulation and coding schemes may be adjusted individually to each SS on a frame-by-frame basis. The uplink channel is based on time division multiple access (TDMA) and dynamic assigned multiple access (DAMA) which is divided into a number of time slots that are assigned for various uses. The uplink channel is controlled by the BS MAC and may vary with time. The channel access mode for the DL channel is time division multiplexing (TDM), with the information for each SS multiplexed onto a single stream of data and received by all SSs within the same sector.

Framing

The PHY specification operates in a framed format. Each frame is divided into a downlink subframe (DL) and uplink subframe (UL). In FDD, the uplink and downlink channels are located on separate frequencies. A fixed duration frame is used for both uplink and downlink transmissions. It allows simultaneous use of both full-duplex SSs (which can receive and transmit simultaneously) and optionally half-duplex SSs (which cannot)

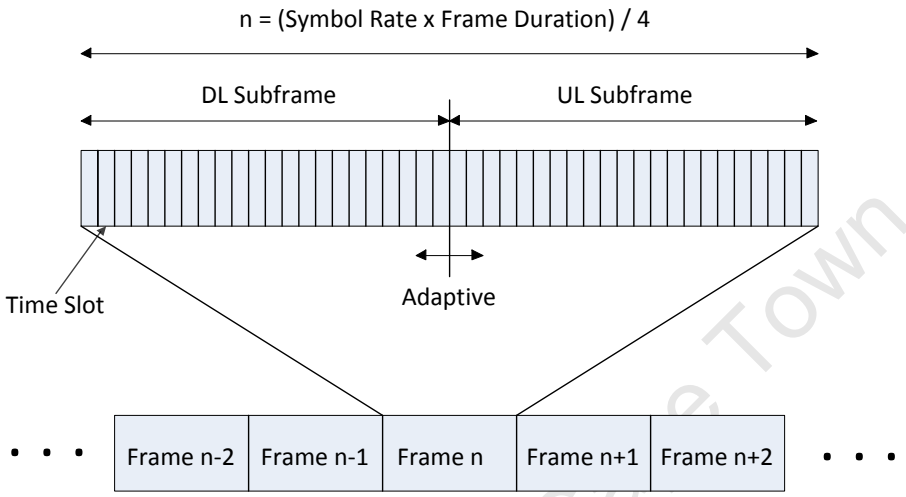

Figure 2.4: TDD Frame Structure

In the case of TDD, the uplink and downlink transmissions occur at different times and usually share the same frequency. A TDD has a fixed duration and contains one downlink and one uplink frequency. A TDD contains one downlink and one uplink subframe with an integer number of physical slots (Figure 2.4), which help to partition the bandwidth easily. TDD is half-duplex. An SS does not transmit and receive at the same time.

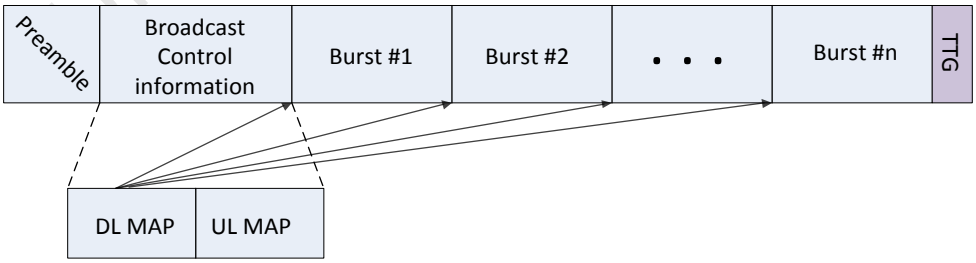

Figure 2.5: The TDD Downlink Subframe Structure

The downlink subframe structure in Figure 2.5 starts with a preamble used by the PHY for synchronization and a frame control section that includes a downlink map (DL-MAP) and an uplink map (UL-MAP) message. Following the frame control section is the TDM portion for

broadcasting data from BS to SSs. A transmit/receive transition gap (TTG) separates the DL subframe from the UL subframe.

The downlink access definition (DL-MAP) is composed of the basic information and several DL-MAP information elements (IE). In a DL-MAP, for each downlink burst, DL-MAP_IE defines the downlink bandwidth allocation of data packets transmitted with the same modulation and coding schemes (MCS) and indicates the start time and channel details of the burst. This type of connection is broadcast to the SSs. A burst may contain multiple connections identifier (CID) field since the packets with the same (MCS) level may belong to different CIDs. The size of a Dl-MAP message depends on the number of downlink data connection.

The uplink access definition (UL-MAP) information element indicates the start time, the duration, and the resource allocation of data burst belonging to the same CID of the same SS including channel details. UL-MAP messages are transmitted in every frame with the most robust MCS level so that each subscriber station will receive information regarding its transmission opportunities even in a bad channel condition. The size of the UL-MAP depends on the number of SSs contained in the UL-MAP.

The uplink subframe structure in Figure 2.6 starts with contention based initial ranging opportunities that allow new SSs to start the initial network entry procedure and request contention opportunities for making bandwidth requests used by SS to transmit the bandwidth request header.

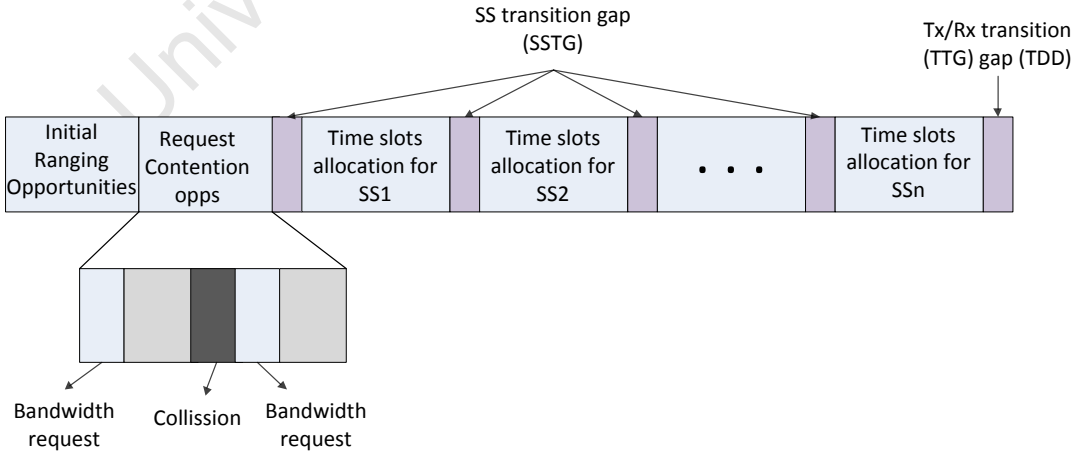

Figure 2.6: The TDD Uplink Subframe Structure

One or more multiple UL bursts are transmitted each from a different SS using the SS

time slot allocation. The SSs transmit data in their respective allocations based on the previously received UL-MAP in the downlink subframe. Subscriber station transition gap (SSTG) separates the transmission into various SSs during the UL subframe. The transmit/receive transition gap (TTG) separates the UL subframe from the DL subframe and allows the SS to switch from the transmit mode to receive mode. Apart from PHY and MAC specifications, 802.16 standards also define QoS enhancements. Some of the enhancements are considered in the next section.

2.2 The IEEE 802.16 QoS Enhancements

Quality of service is the fundamental premise of the IEEE 802.16 MAC architecture. The standard defines some QoS enhancements that can provide support for the different service types in IEEE 802.16. The QoS enhancements are connection and service flows, data unit structure, bandwidth request and allocation schemes, polling mechanisms and different service types. Explanation on connection and service flows, data unit structure and bandwidth request and allocation schemes can be found in Appendix A.

2.2.1 Scheduling Service Types

WiMAX MAC layer uses scheduling services to deliver and handle SDUs and MAC PDUs with different QoS requirements. Scheduling services determine the mechanism the network uses to allocate UL and DL transmission opportunities for the PDUs. WiMAX defines five scheduling services namely Unsolicited Grant Service, real time polling service, extended real-time polling service, non-real time polling service and best effort.

A. Unsolicited Grant Service (UGS)

Unsolicited grant service is designed to support real-time data streams that generate fixed size data packets on periodic basis, such as T1/E1and voice over internet protocol (VoIP) without silence suppression. This service offers constant bandwidth grants periodically according to IEEE 802.16 standard [1]. The polling overhead and latency of bandwidth requests by SSs are eliminated since bandwidth grants are constant on a periodic basis. The QoS parameters associated with the service class are maximum sustained traffic rate (MSTR), maximum latency (ML), and tolerated jitter (TJ).

B. Real time polling service (rtPS)

Real time polling service is designed to support real-time data streams that generate variable size data packets on a periodic basis, such as moving picture expert group (MPEG) video and streaming audio. The service offers real-time, periodic unicast request opportunities that allow SSs to make bandwidth requests and state the grant size required for uplink data transmission. Because of the unicast polling of SSs, increased overhead is incurred by the service type. The service type can also make use of contention slot to make bandwidth request. The QoS parameters associated with the service type are MSTR, minimum reserved traffic rate (MRTR) and ML.

C. Non-real time polling service (nrtPS)

Non-real time polling service is designed to support delay tolerant applications such as file transfer for which minimum reserved traffic rate is required. Unlike rtPS, nrtPS connections are not necessarily polled in a periodic manner but polling must be regular. Multicast and broadcast polling mechanism are also used for the service type. The associated QoS parameters are MSTR and MRTR.

D. Extended real time polling service (ertPS)

Extended real time polling service is designed for transporting real-time data streams with variable data rate (VBR) such as VoIP with silence suppression. It combines the efficiency of UGS and rtPS services. Akin to UGS, the BS provides unicast grants in an unsolicited manner, which sustains the latency of a bandwidth request. Similarly, like rtPS, the bandwidth allocation of ertPS is dynamic since SSs may request changing uplink bandwidth allocation. The associated QoS parameters are MSTR, MRTR, ML and TJ.

E. Best Effort (BE)

Best effort service is designed to support the applications that do not have strict QoS requirements, such as web surfing. For BE connections, all forms of polling are allowed to make bandwidth request. The QoS parameter associated with this class is MSTR.

The background knowledge of composition and operation of IEEE 802.16 networks will enable us to have a better understanding of some literature works that have been done in the areas of admission control and packet scheduling for IEEE 802.16 networks. This leads us to the next section of thorough review of related literature.

2.3 Literature Review

Connection Admission Control (CAC) in IEEE 802.16 Networks

Connection admission control is a vital part in the QoS provisioning process. Some studies have focused on development of CAC since the introduction of IEEE 802.16 Standard. Different algorithms have been employed in making admission decisions for connection requests by a service flow. The simplest of these algorithms is the complete sharing scheme.

Complete sharing (CS) scheme assumes the base station accepts all connections until it runs out of resources. CS is easy to implement and it works efficiently when BS is handling a single type of service. However, CS scheme cannot work efficiently when multiple service types are involved since there would be unfairness in resource allocation. IEEE 802.16 defines five service types that make CS insufficient for WiMAX [23]. Classic approach to CAC in wireless networks assumes allocation of dedicated resources like bandwidth reservation, service degradation to admit new connection request and fixed/dynamic guide channel or threshold to make provision for varying traffics.

Wang et al [24] proposed a CAC scheme that assigns highest priority to UGS flows and aims to maximize bandwidth utilization by using bandwidth borrowing and reduction methods. The UGS flows are allocated a predetermined bandwidth capacity, U of the total bandwidth capacity, B of the network. The value $B - U$ is bandwidth capacity reserved for rtPS and nrtPS connections. They denote b_{ong} as the bandwidth set aside for on-going connections (UGS, rtPS and nrtPS) and b_{ugs} and b_{rtPS} as the bandwidth requirements of new UGS and rtPS connections respectively. The minimum and maximum bandwidth requirements of a new nrtPS are denoted by b_{nrtPS}^{min} and b_{nrtPS}^{max} respectively. Bandwidth reduction is performed when a new rtPS connection is requested and $(b_{ong} + b_{rtPS} > B - U)$. Likewise, bandwidth reduction is performed when a new nrtPS is requested and $(b_{ong} + b_{nrtPS}^{max}. l_{nrtPS}^{n} * \delta)$. The parameter l_{nrtPS}^{n} is denoted as the reduction step and δ is the amount of reduced bandwidth for every reduction step. The reserved bandwidth for each nrtPS connection is given as $(b_{nrtPS}^{max} - l_{nrtPS}^{n}) * \delta$ which satisfies $(b_{nrtPS}^{max} - l_{nrtPS}^{n} \geq b_{nrtPS}^{min})$ and the maximum reduction step is given as $l_{nrtPS}^{max} = (b_{nrtPS}^{max} - b_{nrtPS}^{min})/\delta$. While bandwidth borrowing and reduction ensure that more connections are accepted, the QoS of the ongoing rtPS and nrtPS connections must be guaranteed. In

addition, the authors allocated a predefined value of the total network capacity to UGS connections which could result to bandwidth wastage when not in use. The same approach was used in [25]. In the approach, the authors did not preallocate bandwidth capacity to UGS service as done in [24] but allowed all the service types to fully access the total bandwidth capacity which is an equivalence of complete sharing. For efficient bandwidth utilization and quality of service guarantee, both rtPS and nrtPS connections must be carefully addressed without violating the quality of service of on-going connections.

In [16], the authors proposed a traffic aware Connection Admission Control scheme for broadband mobile systems. The scheme is based on the bandwidth reservation concept, which is basically designed for 'busy hour' of a typical day. The scheme provides higher priority to VoIP calls (UGS connections) compared to other types of traffic (ertPS, rtPS and nrtPS connections) in the network. Like Wang et al [24] a portion, BW_R of the total bandwidth, BW_T is reserved for UGS connections and the restricted bandwidth, $(BW_T - BW_R)$ is provided to the service types of ertPS, rtPS and nrtPS. No bandwidth allocation is considered for BE connections because, its requests are always admitted without bandwidth allocation. The portion of the reserved bandwidth for UGS connections is dynamically changed according to the traffic intensity of the VoIP calls and is given as:

$$(BW_R = [\rho_1 * \beta] * BW_1) \quad \textbf{(2.1)}$$

Where ρ_1 denote the traffic intensity of the UGS connections and BW_1 is the bandwidth needed for each UGS connection, while $\beta \in [0,1]$ denotes the bandwidth reservation factor. In the proposed scheme, the service types are classified into UGS and non-UGS connections. Though this classification simplifies the scheme and assures UGS connections a lower blocking probability, the service types of ertPS, rtPS and nrtPS cannot be classified as one, because their QoS requirements are different. Different service types as well as their different priority level need to be considered for QoS differentiation when designing admission an control scheme.

Admission control for non-preprovisioned service flows in wireless metropolitan area networks was proposed in [26]. The admission control policy uses a guard channel scheme to prioritise handover connections over the new connections by reserving a percentage of the total

bandwidth capacity to the handover connections. The BS scheduler uses class priority to schedule connection requests in the queue. The priority of the queue in a descending order is defined as:

- Handover UGS
- Handover rtPS and handover ertPS
- New originated UGS
- New originated rtPS and new originated ertPS
- Handover nrtPS
- New originated nrtPS
- Handover BE
- New originated BE

A proportional bandwidth-borrowing scheme is achieved by reducing the traffic rate of rtPS and nrtPS to their minimum reserved traffic rates thereby making available more bandwidth to accept more handover and new connections. If after possible bandwidth borrowing has been made, a new connection is blocked if the remaining bandwidth is less than the reserved bandwidth for the handover connection. A handover connection is blocked if there is no capacity to admit the connection. The results obtained show low dropping and blocking probabilities for high prioritized connections while the scheme is highly unfair to nrtPS and BE service. The scheme did not explain how the different service types are differentiated in the admission policy.

Lang et al [27] proposed a joint bandwidth reservation and admission control scheme for IEEE 802.16e networks. In the proposed scheme, two models are defined: model without buffer and model with buffer. In the model without buffer, a min-max optimization problem is formulated to minimize the maximum blocking probability of all classes. Best effort traffic is reserved a fix portion of the total bandwidth since it has no minimum bandwidth requirement. In this model, a new connection is blocked when the numbers of ongoing connections reach the maximum number the network can support. The simulation results when compared with the average allocation scheme, where the network bandwidth capacity is equally divided among all service classes, performs better in terms of call blocking probability and fairness to the service classes. However, the authors did not consider the effect of the proposed scheme among the service classes in terms of blocking probability and fairness in resource allocation. In addition,

the buffering scheme cannot be useful for delay bound connections like UGS, ertPS and rtPS since the service types cannot tolerate delay beyond their delay bounds, otherwise, packets are rendered useless.

Packet Scheduling in IEEE 802.16

The approaches adopted in designing packet schedulers in PMP WiMAX network can be divided into three main categories [28]; queuing-derived approach, optimization-based approach and cross-layer approach. The queuing-derived approach utilizes the scheduling mechanisms such as Weighted Round Robin (WRR), Earliest Deadline First (EDF), Weighted Fair Queuing (WFQ) and Deficit Round Robin (DRR), which are modification of Round Robin (RR), Fair Queuing (FQ) and First-in First-out (FIFO) scheduling mechanisms [29]. Queuing-derived strategy focuses on the queuing aspect of the scheduling problem and try to find the appropriate queuing discipline that meet the QoS requirements of the service classes supported by IEEE 802.16 standards. The hybrid or hierarchical scheduling algorithm combines two or more of these mechanisms. The optimization-based approach tries to formulate and optimize the problems whose objective is to maximize the system performance subject to constraints reflecting in general the QoS requirements of different service classes. The cross-layer scheduling approach considers the optimization of two or three layers and thus improves the system performance but with design complexity.

A layer scheduling based on round robin (RR) discipline is proposed in [30] . The authors argue that one layer scheduling is better than the hierarchical scheduling since a scheduling process can be done through a simple mechanism. The proposed scheme comprises three stages. The first stage, the BS calculates the minimum number of slots required by a connection to guarantee QoS. In the second stage, the unused slots left after the minimum requirement have been satisfied are allocated to rtPS, nrtPS and BE connections in a round robin manner. The third stage specifies the ordering of the slots. In addition, the proposed scheme takes into consideration the overhead resulting from fragmentation and packing. However, the slot ordering is not in accordance with the IEEE 802.16 standard.

Elmabruk et al [31] propose a fair and latency aware uplink scheduler in IEEE 802.16 using customized deficit round robin (CDRR). The proposed scheme is based on grouping different connection types in different queues. A single queue is made for both UGS and unicast

polling and one queue for BE while both rtPS and nrtPS are given a list of queues. The queue list is updated for every frame by adding new queues and removing empty queue from the list. Each queue list is attached with a deficit counter variable to determine the number of requests to be served in the round and this is incremented in every round by a fixed value (quantum). The quantum allocated to a flow is defined as the service the flow should receive during each round robin. The scheme introduced an extra queue to store a set of rtPS requests which are to miss deadlines in the next frame. In the next scheduling cycle the scheduler first serves the UGS and polling queue and then all the requests in the extra queue before missing their deadlines. Once the extra queue becomes empty and there is available bandwidth, the scheduler serves the rtPS and nrtPS lists, using DRR with priority for rtPS, followed by nrtPS. For BE queues the scheduler assigns the remaining bandwidth in first in first out (FIFO) since it has no QoS boundaries. In this work while double priority is given to rtPS requests by having normal queue which is prioritized over nrtPS requests and an extra queue to store the requests which are to miss deadlines in the next frame and the all the requests in the extra queue must be served in the next frame before serving any other queues. This priority scheme will not only starve the nrtPS requests if there are many extra queue requests, but also affect the BE requests which are to share the unused bandwidth. While improving this uplink scheduler scheme, the number of queues will be minimized and nrtPS and BE requests will be protected against starvation. The scheme will be designed to achieve fairness to all service types and ensure that delay bound request do not miss their deadlines. Also ertPS will be introduced so that the five classes introduced by the standard are addressed.

The authors in [6] proposed a packet scheduling scheme for IEEE 802.16 and WiMAX. The proposed scheduler named 2-Tier Ad-Hoc scheduling scheme (2T-AHSS) works in two stages: Ad-Hoc stage and dynamically allocating priority queue (DAPQ) stage (Figure 2.7).

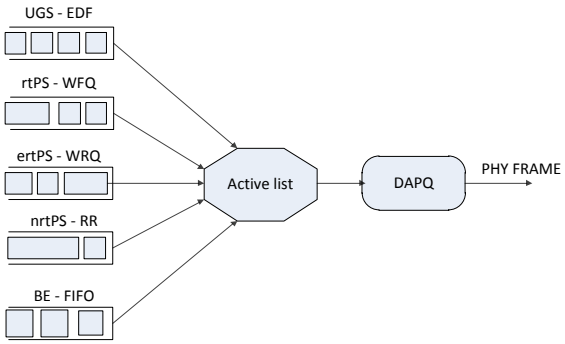

Figure 2.7: 2-Tier Ad-Hoc Scheduler

The scheduler is designed to manage the five scheduling services. The first stage employs five scheduling schemes for the different service types: earliest deadline first (EDF) is used for UGS, weighted fair queuing (WFQ) is used for ertPS and rtPS and round robin (RR) for nrtPS while BE traffic is served with first in first out (FIFO) scheme. The head of the line packet of each queue is moved to the second scheduling stage where DAPQ scheme is used. DAPQ is a modified priority queue (PQ) scheduling algorithm. Unlike PQ which uses strict priority, DAPQ sets a pre-set cap value for each service type and schedules the service flows based on the active list by using the priority policy : UGS > (ertPS, rtPS) > (nrtPS, BE) to deliver the packets to the physical frame. The scheduling scheme consider the five scheduling services and achieves higher throughput when compare with the PQ scheme. However, a pre-set cap for each service flow limits the resource utilization efficiency since a service type cannot receive more the pre-set cap even if other service types are not making use of their resources. In addition, the scheme can lead to starvation of low prioritiy services types since they are not protected against the high priority service types.

In [32], a service flow management strategy for IEEE 802.16 was proposed. The authors design a two layer hierarchical scheduling structure for bandwidth allocation for both uplink and downlink flows. The scheme supports all UGS, rtPS, nrtPS and BE types of service flows. In the first scheduling layer, a Deficit Fair Priority Queue (DFPQ) is used to allocate bandwidth to the multiple service flows (Figure 2.8). The DFPQ visits the non-empty active list maintained in BS and determines the number of the requests in the queue.

A variable Deficit Counter that is incremented by the quantum value each time the scheduler visits a non-empty queue in the active list is used to keep track of the transmitted

packet. Each time a number of bits of a packet are scheduled to be transmitted to the output port; the variable deficit counter is reduced by the number of bits in the transmitted packet. The process is repeated until either the deficit counter is reduced to zero or the queue is empty.

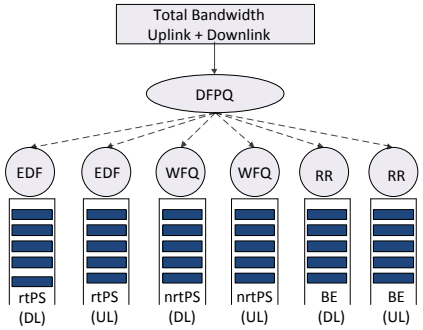

Figure 2.8: Two -Layer Hierarchical Scheduler

For an empty queue, the value of the deficit counter is set to zero. When this condition occurs, the scheduler moves on to serve the next non-empty priority queue. In the second layer scheduling, earliest deadline first (EDF) is used for rtPS, weight fair queue (WFQ) for nrtPS, and round robin (RR) for BE. The bandwidth requirement of UGS is allocated before scheduling, since fixed bandwidth is allocated to the service type. The DFPQ performs better when compared with the strict priority scheme in terms of fairness and throughput. However, the authors did not specify how network capacity would be shared between the uplink and downlink channel. Also, using the same scheduling for both uplink and downlink channel can result in computational complexity. A separate scheduler for uplink and downlink channel in the base station would perform better in terms of resource allocation.

Ganz et al [33] proposed an uplink scheduling algorithm mechanism with CAC that uses hierarchical structure for bandwidth allocation. The overall bandwidth is allocated according to strict priority from UGS, rtPS, nrtPS and to BE. The UGS class has the highest priority and BE the lowest priority. The scheduling of UGS is defined by the 802.16 standard with fixed bandwidth allocation. Earliest deadline first is used for rtPS and weighted fair queue (WFQ) algorithm for nrtPS where packets are based on the weight of the connection (the ratio of the connection's average data rate and the total average data rate). The remaining bandwidth is equally allocated to the BE connections. The proposed scheme is evaluated through experiments, however only rtPS and BE traffic is utilized. In addition, the overall bandwidth allocation based

on strict priority can lead to starvation for the service type with the lowest priority.

Knowing that the uplink scheduler is executed at each time frame and this can be many as four hundred frames per second, simpler solutions become more attractive [34]. In order to guarantee QoS in IEEE 802.16 and support various service types, a simple and efficient scheduler is essential.

Connection admission control has an important role to play especially when combine with packet scheduling. A single CAC algorithm cannot guarantee all the required QoS without the support of packet scheduling. Although the two algorithms can be designed separately, each has to be efficient in managing network resources [35]. Some research work has been conducted to provide QoS in IEEE 802.16 networks by combining CAC and packet scheduling.

In [36], the authors proposed a token bucket based call admission control (CAC) and packet scheduling for IEE 802.16 networks. In the proposed scheme, a bucket size, and a token rate are used to control the traffic injected into the network. A packet is not allowed to transmit until it possesses a token. Before connection establishment with a BS, a traffic flow (UGS, rtPS, nrtPS or BE) sends the QoS parameters to the BS and waits for a response. The rtPS delay requirement parameter is also sent when making admission decision. A limit is set for each service type based on the bandwidth of the network. The scheduling packet algorithm adopted the scheme proposed in [33]. A mathematical model is developed to calculate the queuing delay, loss rate requirement and estimate the appropriate token rate for a traffic flow. After UGS connections have been scheduled, and rtPS, nrtPS and BE are granted their minimum requirements, provided they have not exceeded the set limits, the remaining bandwidth is given to the nrtPS and BE. The scheme when compared with the work done in [33] delivers a high throughput for rtPS. However, the scheme needs an estimation model and the accuracy of the proposed model depends on the accuracy of the estimation.

Borin and Fonseca [34], consider uplink scheduler and admission control for IEEE 802.16 standard. In the admission control scheme, the overhead incurred by the bandwidth request mechanism is put into consideration. The tolerated jitter parameter and the rate used for unicast polling are considered when allocating grants to UGS/ertPS and rtPS/nrtPS connections respectively. The proposed scheduler algorithm utilizes three queues with different priorities. The low priority queue stores BE connections. The intermediate queue stores the bandwidth

requests sent by rtPS and nrtPS connections. The high priority queue stores the periodic grants for the UGS and ertPS connections and unicast request opportunities of rtPS and nrtPS that must be scheduled in the next frame. In addition, bandwidth requests can migrate from intermediate queue to high priority queue when the deadline will expire in two frames ahead so that their QoS can be met. A dual leaky bucket is used for policing to ensure that connections comply with the agreed QoS requirements. When the scheduler is executed, it inserts periodic grants into the high priority queue, checks which rtPS and nrtPS requests should migrate to the high priority queue, distributes the non-allocated bandwidth among the BE connections and schedules the connections. The scheme is not fair to BE connections since it is only non-allocated bandwidth that can be assigned to them. If the traffic of other connections types is heavy, the scheme will not have bandwidth to allocate to the BE connections.

2.4 Chapter Discussion

This chapter has given a brief discussion of different wireless networks and their limitations in meeting the challenges of the growing demand for broadband wireless networks. An overview was presented on IEEE 802.16 WMAN with their MAC and PHY requirements. The QoS service enhancements such as connection flow and service flow, data unit, bandwidth request and grant mechanisms and different IEEE 802.16 scheduling services were extensively discussed. It was identified in the literature review that the QoS of different service types can be efficiently guaranteed by combining CAC and packet scheduling algorithms in IEEE 802.16 networks. This thesis will focus on designing CAC and packet scheduling algorithms that will ensure QoS guarantee for different service types. The knowledge of this chapter would help us to fully understand the requirements of our proposed work.

Chapter 3 Proposed Connection Admission Control and Packet Scheduling

3.1 Introduction

In this chapter, the proposed connection admission control and packet scheduling is presented. In order to provide better understanding, a description of QoS related challenges introduced by the wireless nature of WiMAX and the service requirements by different service types in point-to-multipoint metropolitan area networks are given. IEEE 802.16 was designed to support various types of applications with different QoS requirements.

3.2 Design Requirements

To satisfy QoS guarantee, the proposed connection admission control and packet scheduling scheme is implemented at the medium access control layer of the protocol stack.

The function of connection admission control is to provide the required QoS level and maintain it during the connections. It is not acceptable to let a connection get a service below the minimum service requirement than that contracted. Resource allocation needs to adapt dynamically to the nature of traffic while providing a large dynamic range of throughput to specific users based on their demand. This must be achieved without degrading the overall network performance and without causing starvation of some service flows.

The scheduling algorithms adopt a flexible strategy in allocating time slots to service flows according to their needs and ensures that the allocated resources remain assigned to the service flows for the entire duration of the communication. The proposed scheduling scheme is priority based in order to distribute the available resources efficiently among the various classes of service depending on their QoS requirements. A brief explanation on the proposed connection admission control and packet scheduling operation is given in the next section.

3.3 Connection Admission Control and Packet Scheduling description

The proposed connection admission control is threshold-based. In order to differentiate the IEEE 802.16 connection types, a Quadra-threshold connection admission control is proposed. The function of threshold setting is to limit the number of connections of a certain service type that can be admitted to the network at a particular time. Each service type has an associated threshold limit for controlling the number of connections that can be admitted. The threshold limit is a fraction of the total uplink bandwidth capacity. When a connection request is made by a service type, the admission control algorithm checks if the connection type has not exceeded its set threshold, the connection type is admitted if the threshold limit is not exceeded; otherwise, the connection request is rejected. The threshold limit of a service type is set according to QoS requirement and the assigned priority of the connection type. When a connection request is admitted into the network, the connection makes a bandwidth request to transmit its data. The bandwidth is handled by the uplink scheduler which schedules the connection type that will transmit data in the next uplink subframe. The following section presents the network model of proposed work.

3.4 Network Model

The IEEE 802.16 point-to-multipoint WiMAX network considered in this thesis consists of a base station and a number of subscriber stations positioned within the antenna sector of the base station. Figure 3.1 illustrates the IEEE 802.16 point-to-multipoint WiMAX network considered in this thesis. As shown in Figure 3.1, the components of connection admission control and packet scheduling are at the MAC layer of the protocol stack. The IEEE 802.16 network consists of a base station (BS) and five subscriber stations (SS#1 to SS#5). Connection requests are initiated by the subscriber stations while the base station handles admission control and packet scheduling. Five traffic types are defined in IEEE 802.16 networks with different quality of service requirements. In order to reduce the complexity of the proposed scheme, each of the five subscriber stations generates only one unique type of service; therefore all the five service types are considered. Subscriber station one (SS#1) generates Unsolicited Grant Service (UGS). Subscriber station two (SS#2) generates extended real time polling service (ertPS). Real time polling service originates from subscriber station three (SS#3) and non-real time polling service (nrtPS) from SS#4 while subscriber station five generates best effort (BE) service. The scenario is rich enough to illustrate the problem under consideration and describes the

applicability of this work in a WiMAX network environment. In the next section, the system model of the proposed work is presented.

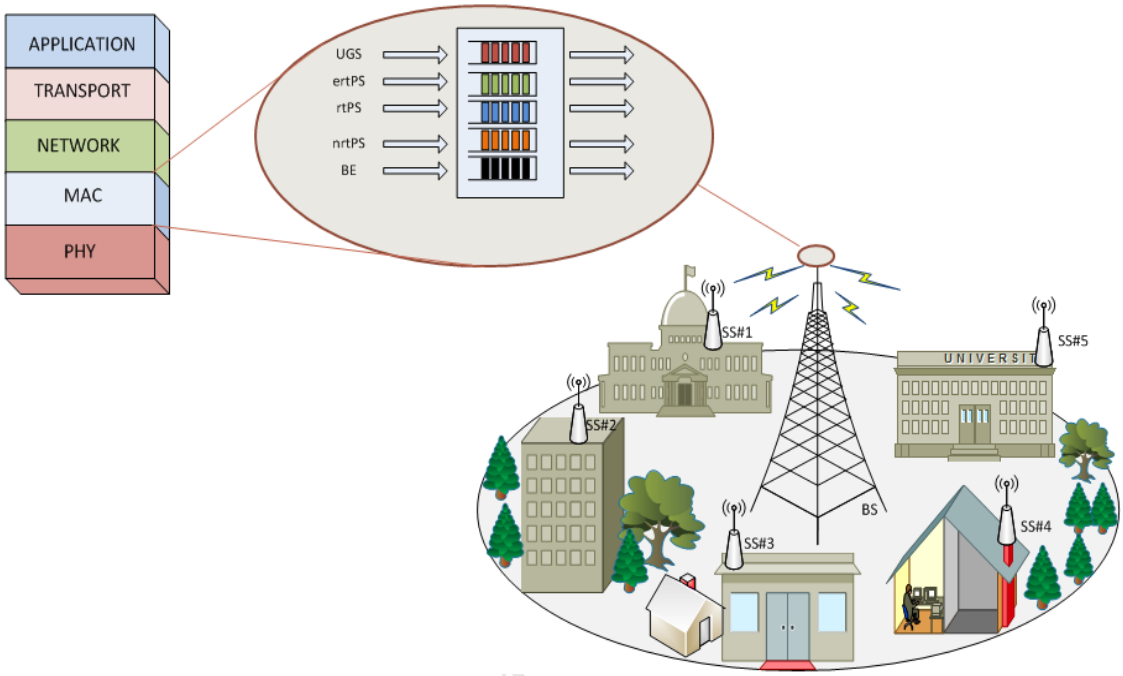

Figure 3.1: Network Model of IEEE 802.16 Networks

3.5 System Model

In order to have a balance between good QoS provisioning and efficient resource utilization, efficient admission control and packet scheduling algorithms are essential. The main objective of CAC in WiMAX networks is to improve the QoS by limiting the number of on-going connections, while a packet scheduler ensures that admitted connections are given their negotiated QoS requirements. Figure 3.2 is the architecture of IEEE 802.16. Generally, CAC operates when a new connection request from a network user is being initiated (Figure 3.2). Before a user can start transmission in the uplink channel, the user must be assured that network resources are available to support the transmission. To be ensured of bandwidth availability, the user makes a connection request through its subscriber station to the base station to which the subscriber station is attached. The CAC in the base station checks whether there is available bandwidth to establish the connection (Figure 3.2). A connection is rejected if the network resources are insufficient to establish the connection otherwise, the connection is admitted.

Admission of a new connection request allows the user to make a bandwidth request which is handled by the scheduler residing in the base station. Based on QoS requirements and priority, the uplink scheduler allocates bandwidth to the connection to send data in the uplink channel. Bandwidth requests are made by newly admitted and on-going rtPS, ertPS, nrtPS and BE connections.

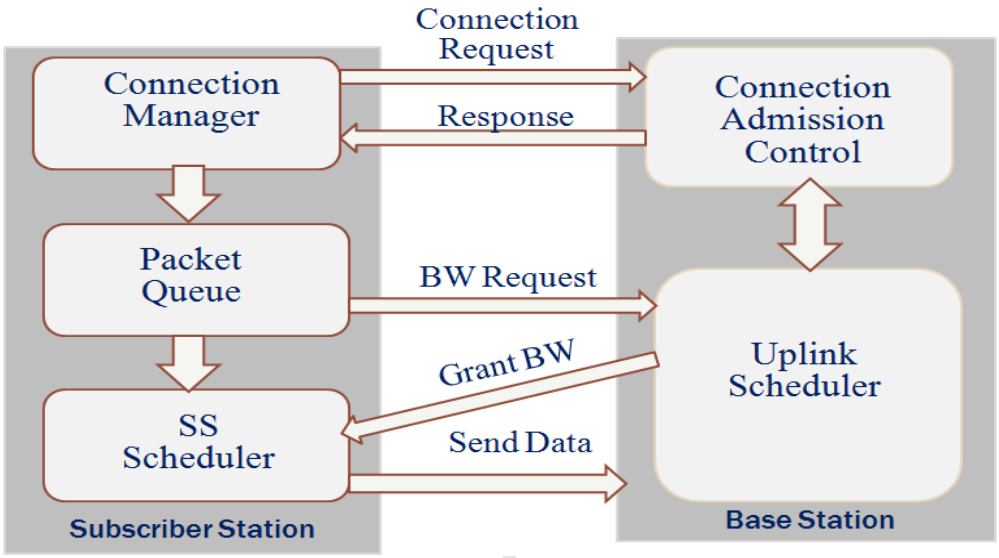

Figure 3.2: Architecture of IEEE 802.16 Networks

A bandwidth request is not made by a UGS connection because it generates constant bit rate data and its bandwidth requirements do not change between connection establishment and termination as defined in IEEE 802.16 standards [4]. Having described the network and system model, the next section presents a detailed explanation of the connection admission control algorithm

3.6 Proposed Connection Admission Control

In this thesis, a novel connection admission control algorithm for reducing the call blocking probability and thereby increasing the resource utilization is proposed. The objective of the admission control is to guarantee users' quality of service requirements. The admission control scheme is focused on the system's ability to accommodate newly arriving connections from network users in terms of the total channel capacity, specified thresholds and associated priority to a service type. Figure 3.3 is the block diagram of the proposed admission control scheme. As shown in Figure 3.3, the connection admission control system consists of three main

components: Connection bandwidth allocation (CBA), Bandwidth Unit Degradation (BRD) and Class Threshold Update, all being coordinated by CAC MANAGER.

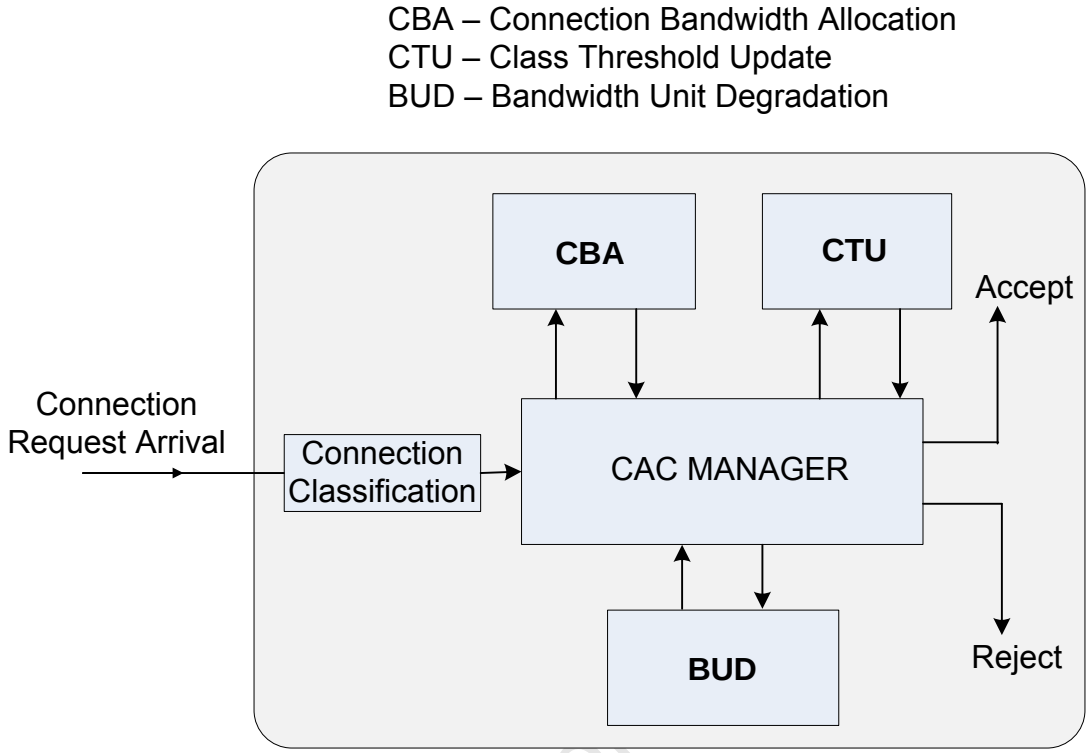

Figure 3.3: Connection Admission Control Framework

As stated earlier, every connection request originates from a subscriber station under the coordination of its base station. Each connection request provides the following parameters: the traffic class, the minimum and maximum traffic rates. Connection requests arriving to the base station are classified according to their service types and QoS requirements by connection classifier. Each of these components is briefly described in the below.

Connection Bandwidth Allocation (CBA)

When a connection request arrives, this component checks for bandwidth availability in the network, the threshold limit of the connection type requesting for admission, estimates the achievable rate that can be offered to the connection and determines if the connection has not violated its admission requirements. This component is triggered by arrival of a new connection into the network. Each connection type has a threshold limit beyond which no connection would be admitted. For example, after the threshold limit set for nrtPS connection is reached, there is no nrtPS connection that would be admitted into the network until resources are free within the

bound of the threshold.

Bandwidth Unit Degradation (BUD)

Bandwidth unit degradation is a vital component of the system. This component is used to reduce the bandwidth of actives connections to the minimum required bandwidth so that more connections can be admitted into the system thereby reducing connection blocking. The component is triggered by a connection request blocking or rejection due to insufficient bandwidth to admit the newly arriving connection. It is noteworthy to state that the bandwidth degradation process is performed only mainly on ongoing nrtPS and rtPS connection. The UGS and ertPS connections are given requested rate while nrtPS and rtPS can be degraded to the minimum reserved rate as defined in 802.16 Standards [4].

Class Threshold Update (CTU)

Class threshold update is responsible for dynamically updating the threshold values of the connection service types based on changes in traffic condition of the connection types in order to react to the changing traffic patterns and to allocate the scarce resources efficiently; the admission threshold of each service types is recalculated periodically. The time between two consecutive threshold updates is referred to as Class Threshold Update Period (CTUP). The time is fixed and threshold update takes place at the end of each CTUP. The time is chosen to be relatively long compared to connection service time so that it would not constitute another processing overhead. The new threshold values are used by CBA in the next CTUP.

In the following subsections, requirements for connection bandwidth, policy for reserving bandwidth, and Quadra-Threshold (QT) technique for sharing bandwidth in the proposed CAC scheme are described.

3.6.1 Connection Bandwidth Requirement

During connection setup, a user sends a connection request informing a base of his service requirement so that quality of service can be guaranteed. Some of the quality of service parameters defined in IEEE 802.16 networks are the minimum reserved traffic rate (MRTR) and the maximum sustained traffic rate (MSTR) [4]. For a UGS connection, there is no MRTR, because of its constant bit rate data stream. An ertPS connection is treated as UGS connection

but can also make request to change its data rate. Both UGS and ertPS are allocated MSTR to meet connection delay requirements. For rtPS and nrtPS connections, offered bandwidths are always between the MRTR and MSTR for efficient bandwidth utilization.

Let the bandwidth requirement of each connection type be represented by a set D. The set D is given as:

$$D = \{ b_u, b_e, b_r, b_n \} \quad \textbf{(3.1)}$$

Where the integers b_u, b_e, b_r, and b_n denote the basic bandwidth unit (bbu) requirements offered to UGS, ertPS, rtPS and nrtPS connections respectively. The offered bandwidth units to rtPS, b_r and nrtPS, b_n are given as:

$$\begin{aligned} b_r^{min} &\leq b_r \leq b_r^{max} \\ b_n^{min} &\leq b_n \leq b_n^{max} \end{aligned} \quad \textbf{(3.2)}$$

Where b_i^{min} and b_i^{max} are the MRTR and MSTR respectively.

3.6.2 Quadra-Threshold (QT) Bandwidth Sharing Scheme

A number of bandwidth allocation schemes for multi-class traffic have been proposed in literature. These schemes can be grouped into the complete sharing (CS) scheme and the bandwidth partitioning (BP) scheme depending on the bandwidth allocation strategy as discussed in section 2 of literature review. In CS scheme, arriving connection requests are accepted based on the condition that network resources are available. No distinction is made between connection types and users of different traffic types are allowed to share all the available recourses. In BP scheme, a certain percentage of network resources are reserved for a certain class or a group of classes. While the CS can be a simple bandwidth reservation policy and be an efficient scheme when all connection requests belong to the same service type they do no guarantee QoS requirement when service types are of priorities. The BP scheme can result to wastage of resources if a connection type is not using the given partition which cannot be accessed by other service type. For an efficient bandwidth allocation strategy and quality of service guarantee, a better bandwidth reservation policy is required, therefore, a multi-threshold bandwidth

reservation referred to Quadra-Threshold bandwidth sharing scheme is proposed.

In the proposed Quadra-threshold bandwidth sharing, each connection type is assigned a bandwidth threshold value according to the priority given to each connection type. The order of threshold priority is given as: $UGS > ertPS > rtPS > nrtPS$. The BE connections are not considered. In 802.16 MAC layer, BE connections get the transmission opportunities only when other service connections do not transmit. Generally, BE connections do have long idle period and data in each transmission is relatively small, especially in the uplink direction. Therefore QoS of BE can be easily satisfied [24], [17].

Let T_v denote the set of threshold values for connection types

$$T_v = \{[t_u, t_e, t_{r,} t_n]: t_n \leq t_r \leq t_e \leq t_u \leq B\} \quad \textbf{(3.3)}$$

Where parameters, t_n, t_r, t_e and t_u are the set threshold limits for nrtPS, rtPS, ertPS and UGS connections and the parameter B, the uplink bandwidth capacity of the WiMAX network respectively. Note that when the parameters $t_u = t_e = t_r = t_n = B$, bandwidth sharing policy becomes complete sharing scheme.

Figure 3.4 illustrates the QT bandwidth reservation policy; all connections request are accepted by the admission controller provided network resources are available to sustain the flows. For quality of service guarantee of connections with high priorities, the admission controller ensures that a connection type does not exceed its set threshold for fair resource allocation. At the start of admission decision epoch when new connection starts soliciting for admission, connection admission requests of all service types are accepted until a set threshold value t_n is reached. Beyond the threshold t_n, new nrtPS connection requests will be blocked. The admission controller would only admit rtPS, ertPS and UGS connection requests until the threshold level t_r after which new rtPS connections will be blocked. After this threshold, only ertPS and UGS connection requests are admitted until the set threshold t_e when the admission controller starts to block new ertPS connection requests. The connection requests of UGS service type are blocked only when there is no capacity in the network to support the connection requests. The QT bandwidth scheme does not only give priority to each connection type by assigning different thresholds but also ensures fairness by allowing a service type to efficiently use the network when other service types are not sending connection requests.

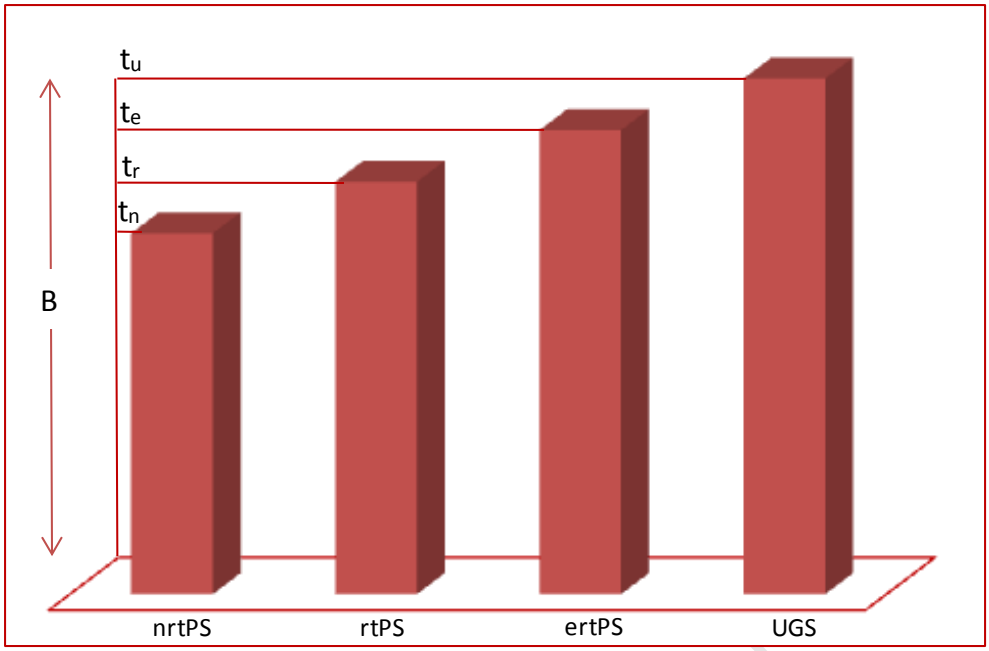

Figure 3.4: Threshold-based bandwidth sharing

3.6.3 Operation of the proposed connection admission control scheme

The operation of the connection admission control scheme is illustrated by a flow chat in Figure 3.5. Arriving connection requests of service type-*i* are classified to the appropriate service class by connection classifier. For each connection type, the admission controller checks if the summation of the required bandwidth of the new connection request b_i and the overall bandwidth offered to the ongoing connections $n_i b_i$ of the four service types is equal to or below the bandwidth capacity of the network. If this condition is true, for nrtPS connection, the connection request is blocked; otherwise, the admission controller checks the second condition. The second condition verifies if the connection has not exceeded its set threshold. Each connection type has a set threshold for quality of service guarantee and differentiation. If the set threshold is exceeded, the connection request is blocked; otherwise, the connection request is accepted and bandwidth is reserved for the connection. This process is performed by the connection bandwidth allocation component of admission controller. When a connection request belonging to nrtPS and rtPS is rejected, this event triggers the bandwidth degradation process which degrades the bandwidth of nrtPS and rtPS to minimum reserved bandwidth requirement so that more connections can be admitted without violating the QoS of the ongoing connections.

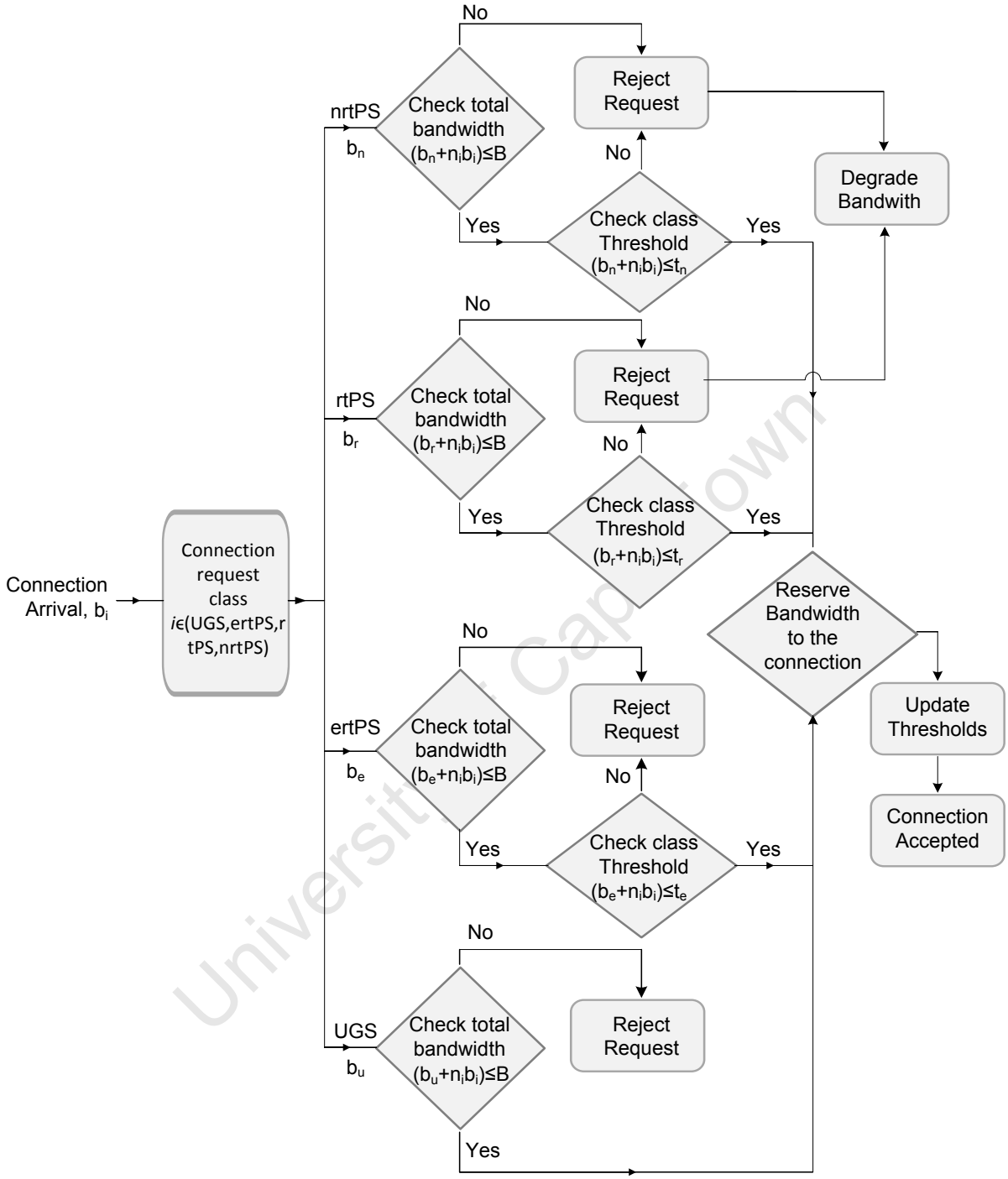

Figure 3.5: Connection Admission Control Flow Chat

The set threshold of UGS service is equal to the bandwidth capacity of the uplink transmission. A UGS connection request is blocked only if the uplink capacity is full. The set threshold is updated periodically according to varying traffic.

3.7 Packet Scheduling

In this section, a priority-based round robin packet scheduler is presented. The scheduler is the modification of processor sharing server [37]. As discussed in section 3.5, after a connection request has been accepted, the base station uplink scheduler maintains virtual queues of bandwidth requests from connection service types based on the amount of bandwidth requested and the amount of bandwidth granted. The scheduler selects the packet of a service type to be transmitted in the next frame according to allocated priority and QoS requirement. According to the connection QoS requirements, the ertPS is assigned the highest priority while the BE service is assigned the lowest priority. The order of priority is given as ertPS>rtPS>nrtPS>BE. Figure 3.6 show the processor sharing round robin scheduler. An unsolicited grant interval is defined for UGS connection in IEEE 802.16 standard where bandwidth is been granted to the admitted UGS connection without making bandwidth requests.

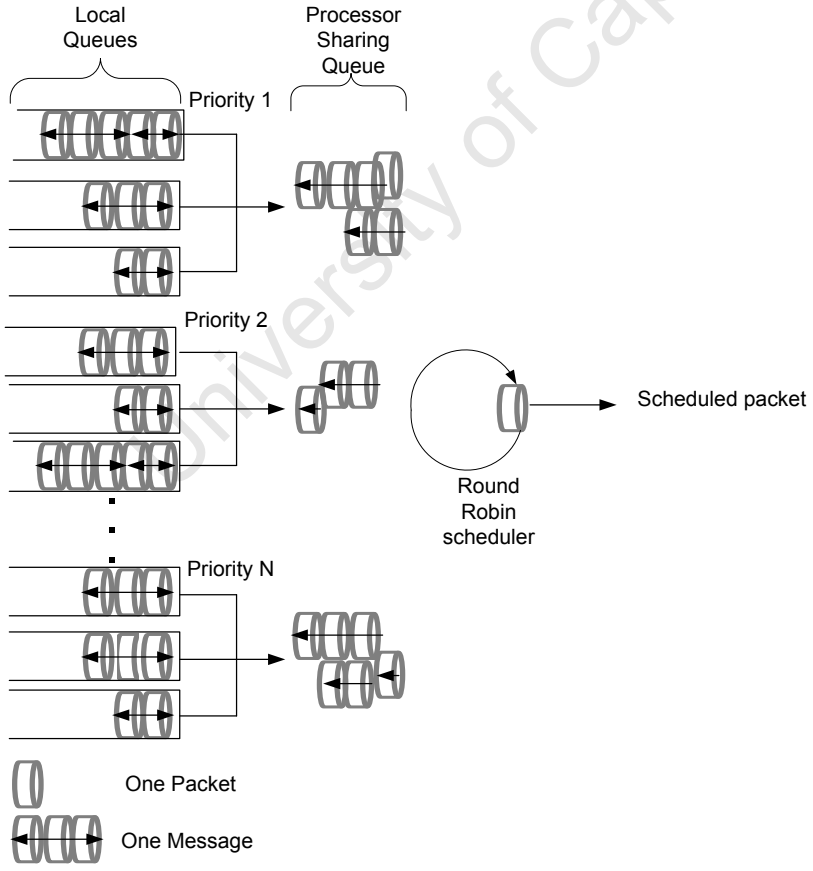

Figure 3.6: Processor Sharing Round Robin Scheduler

As shown in Figure 3.6, each service type of priority $i, \{i = 1, \dots N\}$ has its own queue. When a bandwidth request message arrives, the message is broken into a number of packets and each packet fits into one time slot of an uplink subframe. The scheduler consists of two queues; local queues and processor sharing queue. The round robin scheduler performs round robin processor sharing among local queues by allowing not more than one bandwidth request message from each local queue to be present in the processor sharing queue. Only when an entire message is completed, is its LQ allowed to transfer another message into the processing sharing queue. Local queues of higher priority service types are first served before serving a local queue of lower priority.

3.8 Chapter discussion

This chapter has discussed the requirements and operation of connection admission control and packet scheduling. Different components of the connection admission control framework to provide required QoS are described. A Quadra-threshold bandwidth based connection admission control that ensures connection service type does not exceed the allocated threshold while requesting for connection admission has been presented. This chapter is the basis for analytical framework presented in chapter four which considers algorithm development for connection admission control and packet scheduling.

Chapter 4 Analytical Framework

4.1 Introduction

This chapter deals with the task of developing an efficient analytical framework for performance evaluation of the proposed connection admission control and packet scheduling for WiMAX networks. The need for accurate and fast-computing tools is of primary importance to face the design issue of this promising access technology. The Markov Decision Process is used to model the connection admission control of different service types defined in IEEE 802.16 Standards. It is of important to explain connection arrival to and departure from the connection network.

4.2 Traffic Model

4.2.1 Connection Request Arrival Process

The connection arrival process may be defined in two ways:

(1) By characterizing the number of arrivals per unit time (the arrival rate);

(2) By characterizing the time between successive arrivals (the interarrival time).

We use the variable λ to denote the mean arrival rate. The parameter $1/\lambda$ denotes the mean time between arrivals. The probability distribution of the interarrival time of connection requests is denoted as $A(t)$ and it is given as:

$$A(t) = Prob\{time\ between\ arrivals\ \leq t\}\ and$$

$$1/\lambda = \int_0^\infty t\ dA(t) \qquad \textbf{(4.1)}$$

Where $dA(t)$ is the probability that the interarrival time is between $t\ and\ t + dt$ with interarrival times independently and identically distributed.

4.2.2 Connection Request Service Process

Connection service process can be described by a rate, the number of connections served

per unit time, or by a time, the time required to serve a connection. The parameter μ is used to denote the mean service rate, and hence $1/\mu$ denotes the mean service time. The probability distribution, $B(x)$ of service time is given as:

$$B(x) = Prob\{Service\ time \leq x)\ and$$
$$1/\mu = \int_0^\infty x dB(x) \quad \textbf{(4.2)}$$

Where $dB(x)$ is the probability that the service time is between $x\ and\ x + dx$.

4.2.3 Poisson Arrival and Exponential Service

Poisson process is one of the important models used in queuing theory. Often the arrival process of customers can be described by Poisson process. Poisson process is a vital model when the connections or calls originate from a large population of independent users. Mathematically, the process is described by a counter in which the number of events that occur within a given time period has a Poisson distribution. If λ is the rate at which arrival events occur, and t is the time period over which we observe these events, the parameter of interest is λt which is the number of events that occurred in time t. The number of arrivals $N(t)$ in a finite interval of length t obeys the Poisson (λt) distribution,

$$P\{N(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad \textbf{(4.3)}$$

The interarrival times are independent and obey the $Exp(\lambda)$ distribution;

$$P\{interarrival\ time\ > t\} = e^{-\lambda t} \quad \textbf{(4.4)}$$

Let the set M of connection types in WiMAX networks be given as:

$$M = [UGS(u), ertPS(e), rtPS(r), nrtPS(n)]$$

A connection request of a class type-i for i is an element of M arrives according to Poisson process with mean arrival rates $\lambda_u, \lambda_e, \lambda_r$ and λ_n for UGS, ertPS, rtPS and nrtPS connections respectively. When two or more independent Poisson streams merge, the resulting stream is also a Poisson stream. It follows that multiple arrival process can be merged to

constitute a single arrival process (λ) whose interarrival times are exponentially distributed, as long as the interarrival times of the individual process are exponentially distributed and arrival process independent of each other then the merged stream is given by

$$\lambda = \sum \lambda_i \; \forall i \in M \qquad \textbf{(4.5)}$$

The connection duration (service time) of the service types are exponentially distributed with mean $\frac{1}{\mu_i} \; \forall \; i \in M$. So, the connection service rate (completion rate) is μ_i.

4.3 Markov Decision Process

In Markov chain model, it is possible to represent the behavior of a system, physical or mathematical by describing all the states it may occupy and by indicating how it moves among these states. The system being modeled is assumed to occupy only one state at any moment in time and its evolution is represented by transitions from state to state. These transitions occur instantaneously. That is the transition depends only on its current state and not on its past history, then the system may be represented by a Markov process [38].

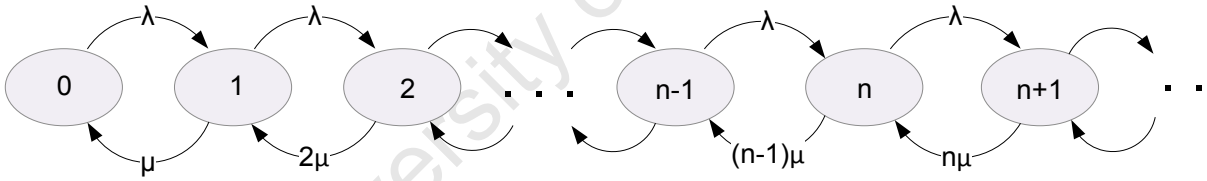

Figure 4.1: 1-Dimentional State Transition of M/M/∞ System

The proposed admission control system in the base station can be modeled as a four dimensional Markov chain where each dimension represents each service type of UGS, ertPS, rtPS and nrtPS. BE connections are always accepted into the network without resource allocation, therefore admission control for the service type is not important [17].

A wireless network with infinite capacity can be modeled as an open queuing network of M/M/∞ queues and the steady state distribution of the number of calls in the network can be given as a Poisson distribution. Figure 4.2 [38] shows the one dimensional state transition rate diagram for an M/M/∞ system. The state transition follows birth and death process where λ is the arrival rate and μ is the service rate. In an M/M/∞ queuing system, we have a situation where

there are always network resources for each arriving connection into the system. If we perform call admission control and limit the numbers of connections admitted into the network to a value which can only be supported by the available network bandwidth, then, the state space of the system is a truncation of M/M/∞ open queuing network. A newly arriving call is served if resources are available; otherwise, the call is blocked. In this queuing model, there is no waiting room for arriving calls. Queuing model has been adopted in some literatures to model call arrival into networks [39], [40].

With the assumptions that connection arrival into the network follows a Poisson process, inter-arrival and service times of connections are exponentially distributed and that the arrival process is independent of each other, the state space of the system can be represented by Markov property and modeled as truncation of 4 independent M/M/∞ queues, therefore has a product form solution where the probability of having n numbers of connections in the network can be calculated.

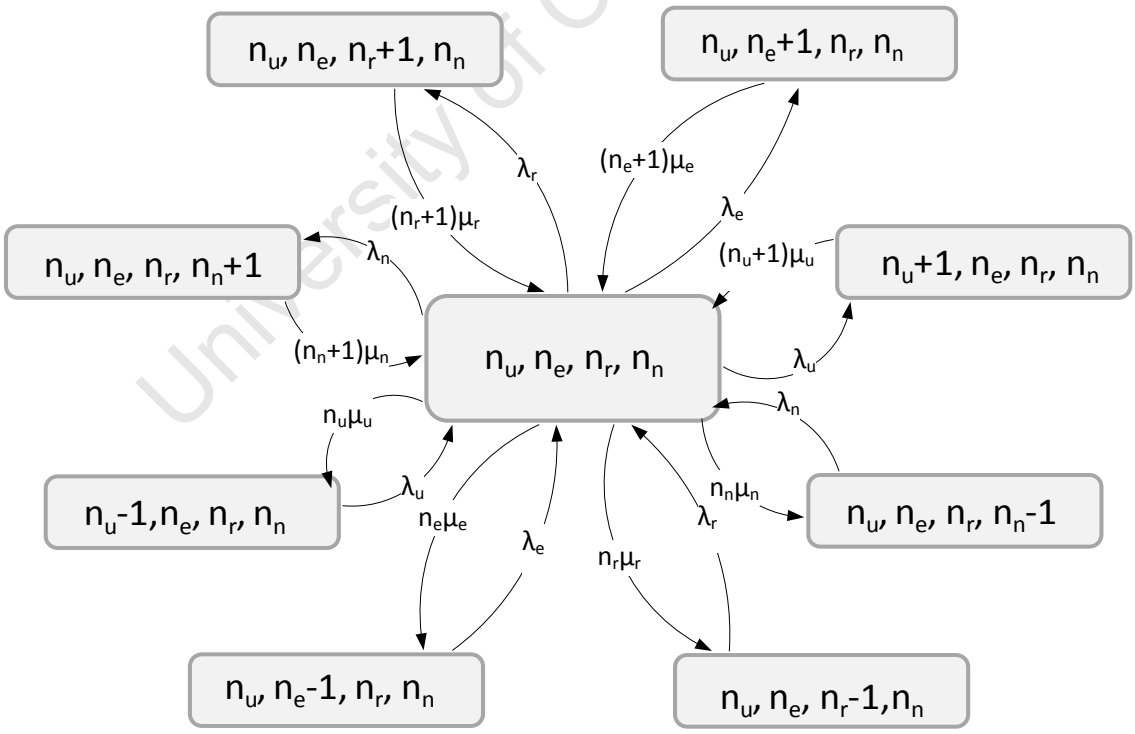

Figure 4.2: 4-Dimentional Markov Model Transition Diagram

Let λ_i and μ_i denote the connection arrival rate and service rate of connection type-*i*,

for $i \in M$. The state of the system in the base station is represented by the vector s. The vector s is given as:

$$s = (n_u, n_e, n_r, n_n) \quad \textbf{(4.6)}$$

Where the non-negative integers $n_u, n_e, n_r \; and \; n_n$ denote the number of UGS, ertPS, rtPS and nrtPS connections in the network respectively. The state s represents the group service of number of connections of service type-i in the base station. The maximum number of connections, N_i of a service type-i that can be present in the network at a given time is given as the ratio of the bandwidth threshold of the service type and its required bandwidth. N_i is given as:

$$N_i = {}^{t_i}/_{b_i} \quad \forall \; i \in M \quad \textbf{(4.7)}$$

For a given state $s = (n_u, n_e, n_r, n_n)$, state transition occurs when a connection request is admitted or when an on-going connection completes. The transition diagram of Figure 4.2 depicts the initial state, the transition state and the transition rate of the Markov chain for the proposed connection admission controller. Given that the initial state of the system is (n_u, n_e, n_r, n_n), when a new UGS connections request arrives to the network and there is sufficient bandwidth to admit the connection, the admission of the request leads to transition from the initial state of the system to state $(n_u + 1, n_e, n_r, n_n)$ with transition rate of λ_u and a transition from this state to the initial state (n_u, n_e, n_r, n_n) depicts that a connection service is completed with transition rate of $(n_u + 1)\mu_u$. Likewise, the state transition from the initial state (n_u, n_e, n_r, n_n) to the state $(n_u - 1, n_e, n_r, n_n)$ shows the completion of a connection with transition rate $n_u \mu_u$ and transition from this state to the initial state depicts the admission of a new connection. The same explanation applies to the other connection types.

The arrival of a new connection of class-i into the network increases the number of the connection types in the network when admitted and the service of a connection of class-i reduces the number of the connection types in the network when completed. The system continues to admit new connections until the threshold limit of the connection type is reached or there is no network capacity to admit the connection.

Let S denote the state space of all possible states. The state of all possible states is given as:

$$S = \{s = (n_u, n_e, n_r, n_n) | (n_n b_n \leq t_n) \wedge (n_r b_r \leq t_r) \wedge (n_e b_e \leq t_e) \wedge (n_u b_u \leq t_u) \wedge \left(\sum_{i \in M} n_i b_i \leq B \; \forall \; i \in M \right)\} \quad \textbf{(4.8)}$$

From equation (4.8), the state space of all possible states in the system is the state such that the total bandwidth of all ongoing connections of nrtPS must be less than or equal to the set threshold, t_n for the service class. Likewise, the total bandwidth of the ongoing connections of rtPS must be less than or equal to the set threshold for the service class. Similarly, the total bandwidths of ertPS and UGS ongoing connections must be less than or equal to their set thresholds. In addition, the bandwidth of the ongoing connections of all the service types must be less than or equal to the bandwidth capacity of the network.

Let ρ_i denote the load generated by a connection type-i. The load generated is given as:

$$\rho_i = \frac{\lambda_i}{\mu_i} \quad \textbf{(4.9)}$$

Let $P(h)$ denote the steady state probability that the system is in state h. State h is the state of the system in which the combination of number of connections in each service class can be simultaneously supported by the capacity of the network without violating the QoS requirement of the connections. The steady state probability that the system is in state h is given as:

$$P(h) = \frac{1}{\pi_0} \prod_{i \in M} \frac{\rho_i^{n_i}}{n_i} \; \forall \; i \in M$$

$$\pi_0 = \sum_{h \in S} \prod_{i \in M} \frac{\rho_i^{n_i}}{n_i} \; \forall \; i \in M \quad \textbf{(4.10)}$$

The parameter π_0 is the normalization constant.

From the steady state solution of the Markov model, performance measures of interest can be determined by summing up appropriate state probabilities.

4.4 Bandwidth Degradation

When a new nrtPS and rtPS connection request is blocked, a bandwidth borrowing process is performed by degrading the bandwidth of ongoing nrtPS and rtPS connections to their minimum bandwidth requirement, thereby more connection requests can be admitted. The number of ongoing nrtPS and rtPS connections has earlier be given as n_n and n_r respectively.

The reserved bandwidth for each nrtPS connection is b_n^j $(0 \leq j \leq n_n)$.

The reserved bandwidth for each rtPS connection is b_r^j $(0 \leq j \leq n_r)$.

The amount of bandwidth that can be borrowed from each nrtPS connection after degradation is $(b_n^j - b_n^{min})$ and the total bandwidth that can be borrowed from nrtPS connections is given as $\sum_{j=1}^{n_n}(b_n^j - b_n^{min})$.

A new connection request b_k is accepted if $b_k + n_i b_i + \sum_{j=1}^{n_n}(b_n^j - b_n^{min}) \leq t_k$ and the request is reserved a minimum required bandwidth, otherwise, the connection is rejected. The parameter b_k is connection request belonging to nrtPS and rtPS and $n_i b_i$ is the bandwidth occupied by ongoing connections of the four service types while t_k is the admission threshold of service type-k. After a possible degradation has been made on nrtPS connections and a new connection request is rejected, a degradation process is also performed on rtPS connections. The amount of bandwidth borrowed from each rtPS connection is $(b_r^j - b_r^{min})$. The total bandwidth borrowed from rtPS connection is $\sum_{j=1}^{n_n}(b_n^j - b_n^{min})$. For a new connection request b_k, if $b_k + n_i b_i + \sum_{j=1}^{n_n}(b_n^j - b_n^{min}) + \sum_{j=1}^{n_n}(b_n^j - b_n^{min}) \leq t_k$, the connection request is accepted and reserved the minimum required bandwidth, otherwise, the request is rejected. At this point, all possible bandwidth degradation process has been done and no connection requests of nrtPS and rtPS services can be admitted anymore. Furthermore, it is good to state that while degradation process is performed only on nrtPS and rtPS connections, the borrowed bandwidth can be used to accept connection request belonging to any of the service types. Because, connection requests of all service types are admitted into the system until a set admission threshold limit is reached for a

particular service type. As we shall see later, the benefit of bandwidth obtained from degradation process is enjoyed by all connection types.

4.5 Class Threshold Determination

In this subsection, an expression is formulated for determining the threshold for each of the four service types. Let the total bandwidth requests of a connection type *i* for *i* is an element of M be limited by some function *f* of the total uplink bandwidth capacity, B; connection type *i* with a total bandwidth requests below the control threshold T can have its bandwidth requests admitted.

At time t, let $T_i(t)$ be the control threshold of connection type-*i*. The control threshold is given as:

$$T_i(t) = f(B) \tag{4.11}$$

The simplest approach is to set the control threshold to a multiple $\propto$ of the total uplink bandwidth capacity. Therefore, it is given as

$$T_i = \propto_i B \tag{4.12}$$

Where $\propto_i$ is the threshold determinant of connection type *i* and it is given as:

$$\begin{aligned} \propto_i &= [B\_R_i] * \left[Q_f * P_{W\,i}\right] \forall\, i \in M \\ B\_R_i &= \frac{H - b_i}{H}, \qquad H = \sum_{i \in M} b_i \;\; \forall i \in M \\ Q_f &= \frac{H^2}{4\sum_{i \in M}(b_i)^2} \;\; \forall\, i \in M \end{aligned} \tag{4.13}$$

The threshold of each connection type is given as:

$$\begin{aligned} t_u &= \propto_u B \\ t_e &= \propto_e B \\ t_r &= \propto_r B \end{aligned} \tag{4.14}$$

$$t_n = \propto_n B$$

The parameter B_R_i denotes the bandwidth ratio factor, Q_f is the fairness quotient factor derived from Jain's fairness index. Since the connection type with small basic bandwidth unit (bbu) requirement will have definitely have low blocking probability, a predefined traffic priority weight denoted as P_{W_i} is used to protect the connections with big bbu from small bbu connections. Equation (7) is bounded by the condition given as $0.7 \leq \propto_i \leq 1$. It is important to state that the threshold determinant of UGS connections is always set to 1 so that the total uplink bandwidth is accessed by the connection type.

4.6 Packet Scheduler

The base station uplink scheduler is modeled as a processor sharing system where the bandwidth request message delay is modeled as the time the packet spends in a local queue and the processor queue until the point of complete departure.

The processor sharing (PS) round robin scheduler is a discrete-time model where time is divided into equal length called time slots. Bandwidth request messages arriving at the local queues consist of an integral number of service times of a single service slot. It is assumed that the number of priority p messages arriving at a local queue within a time slot are independent and identically distributed (i.i.d) and are also independent of arrivals into other local queues (see Figure 3.6), the number of packets contained in a message (the message length) are discrete i.i.d. for each priority, and message transmission can only be interrupted by messages from higher priorities or from other connections of the same priority after the current packet is completely transmitted, i.e. until the end of this slot time.

The mean delay $D_p(k)$ (in unit of time slots) of a priority p message of length k packets is given as:

$$D_p(k) = L_p + S_p(k) \qquad \textbf{(4.15)}$$

Where L_p is the mean time of priority p message in the local queue and $S_p(k)$ is the mean time of priority p message consisting of at least k packets, spends in the PS queue to complete the service of k packets. The priority $p = 1,2, \dots, P$ where 1 is the highest priority.

Let a random variable n_p represents the number of priority p message arrivals within a time-slot to any priority-p local queue. The mean of n_p is denoted by $\overline{n}_p$.

Let the random variable e_p be the priority p message length with mean $\bar{e}_p$. Since a packet transmission requires a time-slot, e_p also represents the message transmission time in units of one- time slot. Let $C_{n,p}^2$ and $C_{e,p}^2$ represent the squared coefficient of variation of n_p and e_p respectively. The total arrival rate of priority p traffic $\lambda_p = M_p\,\overline{n}_p$ and the total traffic load of priority-p traffic, $\rho_p = \lambda_p \bar{e}_p$. Let $\varepsilon_p = \sum_1^p \rho_i$. The parameter M_p is the number of connections in a local queue of p service class. According to [37] we have

$$D_p(k) = \frac{v_p + \sum_{i=1}^{p} v_i/(1-\varepsilon_p)}{2(1-\varepsilon_{p-1})} + \frac{k - [\bar{e}_p(1+C_{e,p}^2)+1]/2}{1-\varepsilon_{p-1}-\frac{M_p-1}{M_p}\rho_p} + \frac{1}{2} \qquad \textbf{(4.16)}$$

$$v_p = \rho_p \bar{e}_p \left(C_{e,p}^2 + \lambda_p \, C_{n,p}^2 / M_p\right) \qquad \textbf{(4.17)}$$

The overall mean priority p message delay is simply given as: $D_p(\bar{e}_p)$.

4.7 Performance Metrics

In this section the performance metrics used to validate the performance of the proposed scheme are considered. As discussed in the previous section, the performance metric for the processor sharing round robin scheduler is message delay given by equation (4.16). The other performance metrics are the metrics considered for the connection admission control scheme.

4.7.1 New Connection Blocking Probabilities

The probability that a new connection arrives into the system and finds that there are no available channels to service it based on the proposed connection admission control policy is known as the new connection blocking probability. A new connection is blocked if the cutoff threshold for accepting the connection has been reached. Therefore, the blocking probability of a

new connection is the sum of the probabilities of state in which the new connection cutoff threshold is reached and exceeded.

4.7.1.1 Blocking probabilities of a new nrtPS connection

Let S_n denote the set of states in which a new nrtPS connection is blocked in the system. A new nrtPS connection request is blocked when the set threshold is reached or there is no more bandwidth capacity in the network to support the connection within its threshold limit.

The set of states S_n is given as:

$$S_n = \{h \in S : (b_n + \sum_{i \in M} n_i b_i) \geq t_n \; i \in M \tag{4.18}$$

The blocking probability of a new nrtPS connection P_n in the system is as:

$$P_n = \sum_{h \in S_n} P(h) \tag{4.19}$$

4.7.1.2 Blocking Probability of a new rtPS connection

Let S_r denote the set of states in which a new rtPS connection is blocked in the system. A new nrtPS connection is blocked when the set threshold is reached or the total bandwidth capacity of the network is used up.

The set of states S_r is given as:

$$S_r = \{h \in S : (b_r + \sum_{i \in M} n_i b_i) > t_r \; \forall \; i \in M \tag{4.20}$$

The blocking probability of a new rtPS connection P_r in the system is given as:

$$P_r = \sum_{h \in S_r} P(h) \tag{4.21}$$

4.7.1.3 Blocking Probability of a new ertPS connection

Let S_e denote the set of states in which a new ertPS connection is blocked in the system. A new ertPS connection is blocked if when the set threshold is reached or there is no more network capacity to admit the connection within its threshold limit.

The set of states S_e is given as:

$$S_e = \{h \in S : (b_e + \sum_{i \in M} n_i b_i) > t_e \ \forall \ i \in M \quad \textbf{(4.22)}$$

The blocking probability of a new ertPS connection P_e in the system is given as:

$$P_e = \sum_{h \in S_e} P(h) \quad \textbf{(4.23)}$$

4.7.1.4 Blocking probability of a new UGS Connection

Let S_u denote the set of states in which a new UGS connection is blocked in the system. A new UGS connection is blocked if the there is no bandwidth capacity in the network. Note that the set threshold for UGS connections is equal to the bandwidth capacity of the network.

The set of states S_u is given as:

$$S_u = \{h \epsilon S : (b_u + \sum_{i \epsilon M} n_i \, b_i) > B \ \forall i \in M \quad \textbf{(4.24)}$$

The blocking probability of a new UGS connection P_u in the system is given as:

$$P_u = \sum_{h \in S_u} P(h) \quad \textbf{(4.25)}$$

4.7.2 Connection Throughput

Connection throughput is defined as the effective arrival rate, the rate at which connection requests enter the system. In a system where a connection request can be blocked, the throughput cannot be defined as connection arrival rate because not all connection requests are admitted. The probability that an arriving connection request is blocked in the system is P_i for $i \in M$, where M is the set of UGS, ertPS, rtPS and nrtPS service types. The probability that the system is not full and an arriving connection requests is accepted into the system is $1 - P_i$. Thus the throughput T_i of a connection type-*i* is given as:

$$T_i = \lambda(1 - P_i) \quad \textbf{(4.26)}$$

4.8 Implementation Approach

This section describes the approach followed in the development of the proposed scheme and the various tools employed.

4.8.1 Software

In this thesis, the MATLAB software tool [41] is used under MS-Windows environment for the implementation of the proposed scheme.

MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numeric computation.

Programming and developing algorithms is faster with MATLAB than with traditional languages, such as C, C++, and FORTRAN. In addition, MATLAB supports interactive development without the need to perform low-level administrative tasks, such as declaring variables and allocating memory. Thousands of engineering and mathematical functions are available, eliminating the need to code and test them by oneself. At the same time, MATLAB provides all the features of a traditional programming language, including arithmetic operators, flow control, data structures, data types, object-oriented programming, and debugging features.

4.8.2 Hardware

The algorithm development with debugging is performed on a Pentium E5400 computer running MS-Windows. The hardware specification is provided in appendix C.

4.8.3 Implementation Steps

The implementation of connection admission control and packet scheduling algorithms consist of three stages namely: Input, Analysis and Output. The block diagram of Figure 4.3

shows the implementation stages.

The first stage involves the supply of input parameters. In this stage, the parameters are generated and fed into the next stage.

The second stage is the stage where computation and analysis take place according to the proposed algorithm developed into a program script. This stage does the analysis of computing the different transition probability and the steady state probability.

The third stage computes the outcome of the analysis and displays the result. It is noteworthy to state that the simulation code for the three stages is written by using a single MATLAB file (M-file). The code is implemented in the MATLAB environment to generate the output results.

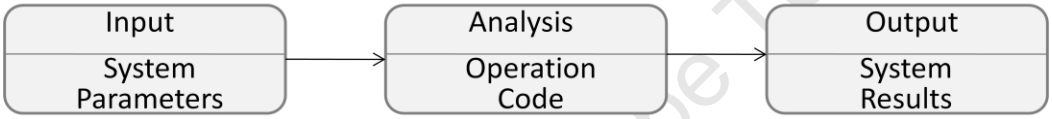

Figure 4.3: Algorithm Implementation Steps

4.9 Chapter Discussion

This chapter has discussed the analytical framework of the proposed admission control and packet scheduling. A brief explanation of the traffic model of the proposed work was presented. Markov decision process has been used to model connection admission control of connection request into the WiMAX network. The different components of the admission control framework for resource allocation are thoroughly addressed. A processor sharing round robin scheduler has been presented. The performance metrics for validation of the proposed work were explained. In addition, the implementation stages and the tools involved were briefly discussed. In the following chapter the performance results are presented with thorough analysis to show the efficiency of the proposed scheme.

Chapter 5 Performance and Result analysis

In this chapter, the performance of the proposed scheme is evaluated. A simulation program is developed and it is implemented by using MATLAB [41]. For connection admission control performance analysis, connection requests of each service type arrive to the base station according to the Poisson process. The connection arrival rate is the same for each service type and ranges from 2-20 connections per second. The limiting threshold of each connection type is calculated using equations (4.13) and (4.14). Other simulation parameters are provided in Table 5.1. The Table 5.1 shows the parameters used in the performance evaluation.

Table 5.1: Parameters used for Performance Evaluation [42]

Parameter Settings			
Service Class	**Min bandwidth (bbu)**	**Max Bandwidth (bbu)**	**Service Class Threshold**
UGS		1	100
ertPS		2	88
rtPS	3	5	82
nrtPS	3	5	80

The performance of the connection admission control is evaluated under three scenarios.

5.1 First Scenario

In the first scenario, the proposed connection admission control scheme denoted as QT is compared with a bandwidth partitioning scheme (PS) [42] and Non-CAC scheme. In the bandwidth partitioning scheme (PS), the uplink bandwidth capacity is partitioned into four parts and each part can only be used by a unique connection type. This method has been used by authors in [24] and [16] to partition the uplink capacity into two parts and each part can only be accessed by a designated group of connection types. In the Non-CAC, the number of connection requests of each service type admitted to the system is not limited to the capacity that can be handled under the controlled threshold of each service type, thereby affecting the performance of the system.

5.2 Second scenario

In this case the effect of degrading the bandwidth requirement of nrtPS connections on all the connection types is examined. Firstly the connection admission control scheme is allowed to offer nrtPS connection request the maximum required bandwidth unit; in this case maximum bandwidth unit is the admission criteria. In the second phase each connection request of nrtPS is degraded to the average required bandwidth unit where the bandwidth average requirement is the admission criteria. The average bandwidth requirement is the mean of minimum and maximum bandwidth requirements. Lastly the bandwidth requirement is reduced to the minimum bandwidth unit which is taken as admission criteria. The offered bandwidth of an nrtPS connection cannot be less than the minimum required bandwidth which [4].

5.3 Third scenario

In this scenario we consider the throughput of each connection type under degradation process of nrtPS connection requests.

5.4 Results

5.4.1 Connection Admission Control

In this section, the analysis of the performance results of the proposed connection admission control is thoroughly explained. The performance result of the first scenario in which the developed algorithm is compared to other scheme is first considered.

5.4.1.1 Scenario 1

In this scenario, the average bandwidth requirement of nrtPS and the minimum bandwidth requirement of rtPS are used in the simulation work.

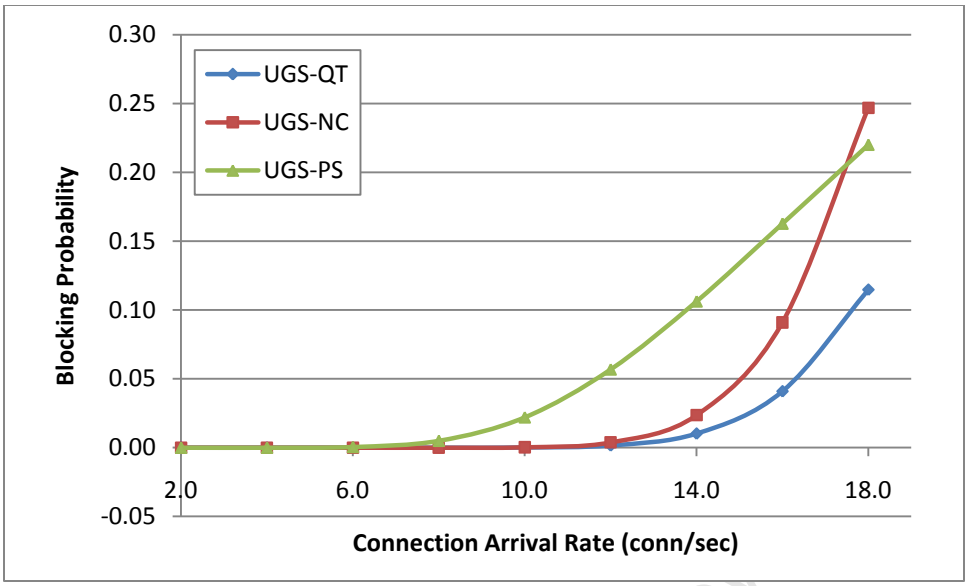

Figure 5.1: Blocking Probability of UGS Connections with Different Schemes

Figure 5.1 shows the blocking probability (BP) of UGS connections against connection arrival rate. The developed scheme, UGS-QT when compared to the scheme without admission control (UGS-NC) and complete partitioning scheme (PS) performs better. The UGS-QT achieves the lowest blocking probability of 0.11 when compared to UGS-PS of 0.22 and UGS-NC of 0.25 blocking probability. The UGS-NC performs better than UGS-PS until the blocking probability of 0.20 when the UGS-NC scheme starts to block more connection requests due to requests flooding the system since there is no admission control. The UGS-PS scheme admits connections requests until the allocated bandwidth portion is fully occupied. The unused bandwidth of other connection types cannot be used by the UGS-PS scheme because; the scheme can only function within the assigned partition. Unlike UGS-PS scheme, the UGS-QT scheme has complete access to the total bandwidth through its threshold setting, therefore, when other connection types do not solicit for bandwidth usage, the UGS connections can make use of the entire bandwidth and lowest blocking probability is achieved.

Figure 5.2 depicts the blocking probability of rtPS connections against arrival rate under different admission control schemes. The developed scheme, rtPS-QT maintains zero blocking probability until the 12th arrival rate when an increase in blocking probability (BP) starts and increases to 0.20 after the 18th arrival rate. Compared to the scheme without CAC (rtPS-NC) and partitioning scheme (rtPS-PS), the rtPS-QT achieves the lowest BP followed by rtPS-NC while rtPS-PS suffers highest BP of 0.41. The BP of rtPS-PS increases drastically, because the

allocated bandwidth for rtPS connections is confined within the given partition which cannot be exceeded even when other service types are not making use of the allocated bandwidth in their partitions.

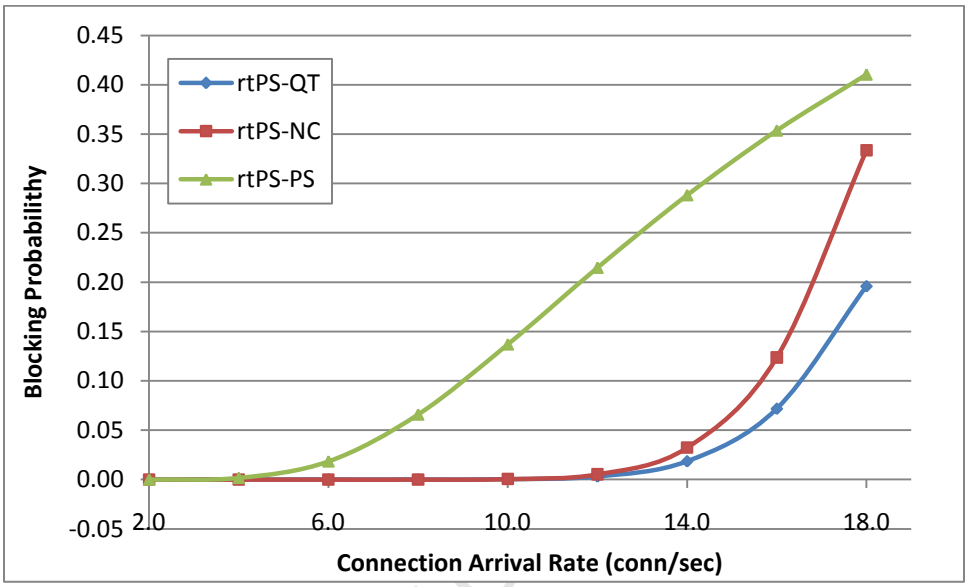

Figure 5.2: Blocking Probability of rtPS Connections with Different Schemes

The rtPS-QT is able to take the advantage of unused bandwidth to admit more connection requests within the set threshold if other connection types are not sending connection request thereby reduces the blocking probability of the connection.

In Figure 5.3, we have the result of blocking probability of nrtPS connections compared under the three schemes. It is noteworthy to state that nrtPS connections generally do suffer the highest blocking probability because of lowest priority assigned to the connection type. Nevertheless, the nrtPS-QT scheme still performs better than the other schemes through the threshold setting which takes the advantage of absence of other connection type requests and utilizes the uplink bandwidth until the set threshold is reached.

The blocking probability of nrtPS-NC scheme increases from 0 at 4th arrival rate to 0.46 at 18th arrival rate having the highest blocking probability while the blocking probability of nrtPS-NC scheme increases from 0 at 12th arrival rate to 0.4 at 18th arrival rate with better performance over the nrtPS-PS partitioning scheme.

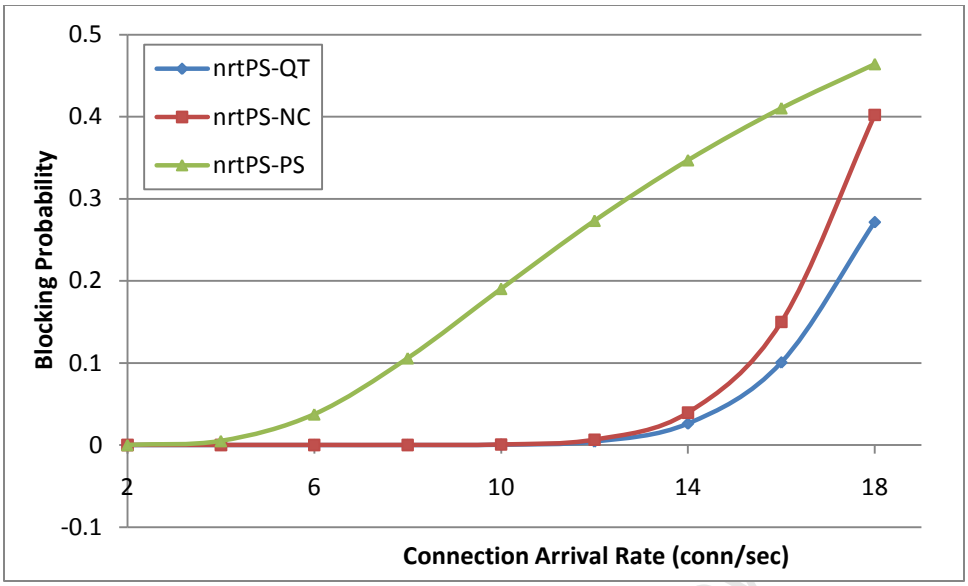

Figure 5.3: Blocking Probability of nrtPS Connections with Different Schemes

The performance of ertPS connection is presented in Figure 5.4. As can be seen, the ertPS-QT maintains zero blocking probability until 12th arrival rate when the blocking probability starts to increase and increases to 0.2 at 18th arrival rate. This at the same time performs better than the ertPS-NC and ertPS-PS schemes by admitting more ertPS connection request with the lowest blocking probability.

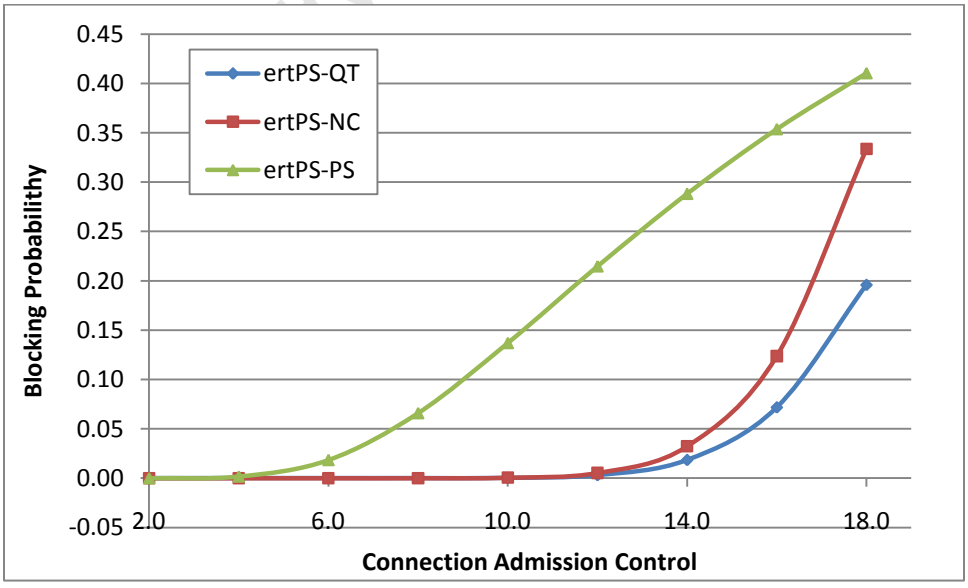

Figure 5.4: Blocking Probability of ertPS with Different Schemes

5.4.1.2 Scenario 2

In this section the blocking probabilities of the four service types against connection arrival rate are considered under different degradation of bandwidth requirement of nrtPS connection.

Figure 5.5 shows the behavior of the four service types when the nrtPS connection is offered the maximum bandwidth requirement. The connections of nrtPS suffer the highest blocking probability increase of 0 at 12th arrival rate to 0.62 at 18th arrival rate because at maximum bandwidth requirement the number of nrtPS connections that is admitted is reduced due to high bandwidth usage of each connection, the set threshold is quickly reached and other connection requests are blocked thereby resulting in a high blocking probability of the connection type. Though the UGS connections achieve the lowest blocking probability followed by rtPS and nrtPS, they are affected by the high bandwidth requirements of nrtPS connection which takes the advantage of the situation where there are few connection requests of other service types and utilizes the system bandwidth requirements. We shall see in Figure 5.15 bandwidth requirement affects the connection throughput.

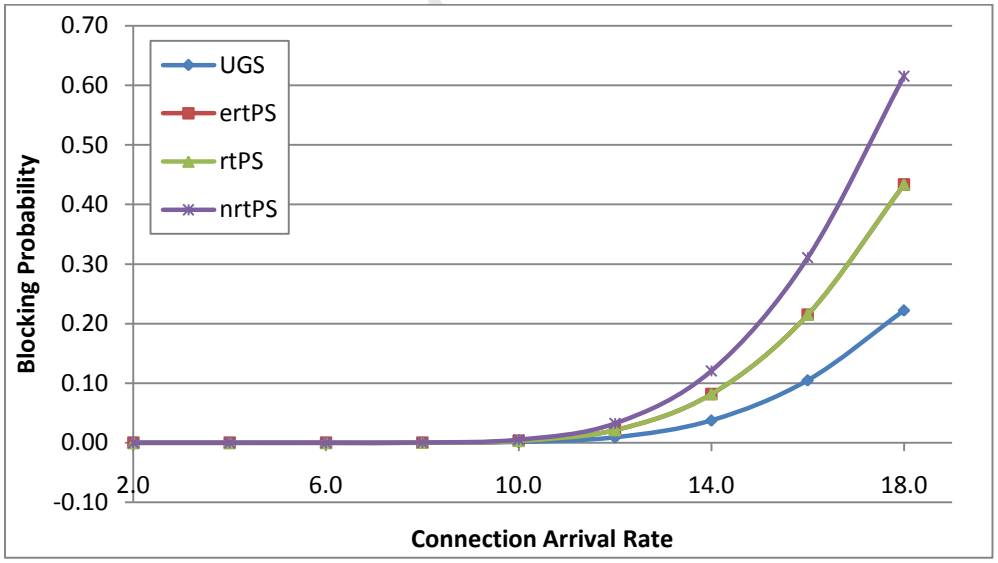

Figure 5.5: Blocking Probability of Connection types under Maximum bbu of nrtPS Connection

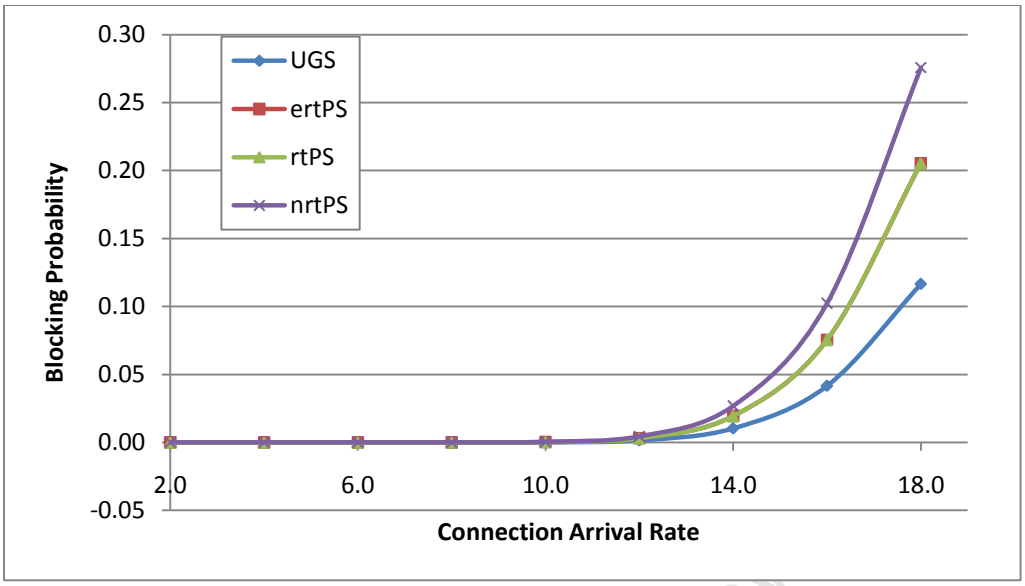

Figure 5.6: Blocking Probability of Connection types under Average bbu of nrtPS connection

In Figure 5.6, the blocking probability of the service types are reduced due to the fact that after degrading the offered maximum bandwidth of the ongoing nrtPS connection to their average rate to admit more connection requests into the service, the newly admitted connections are offered average bandwidth requirements. This in turn reduces the blocking probability of the connection types. As stated in section 4.4, the degraded bandwidth can be used to admit connection requests belonging to any of the service types. Comparing Figure 5.5 to Figure 5.6, it is seen that lower blocking probability is achieved in Figure 5.6. This occurs because, new nrtPS connection requests are offered average bandwidth which makes allowance for more connection requests to be admitted thereby reducing the blocking probabilities of the connections.

The connection blocking probability when the bandwidth unit requirement of nrtPS connections is limited to the minimum required bandwidth is given in Figure 5.7. In this case, lowest blocking probability is recorded. More connections are admitted due to the fact that the offered bandwidth to each nrtPS connection is a minimum bandwidth requirement. It is noteworthy to state that connection types cannot exceed their set thresholds and therefore the connection type with higher threshold suffers lower blocking probability and the one lower threshold suffer higher blocking probability. This is good QoS differentiation.

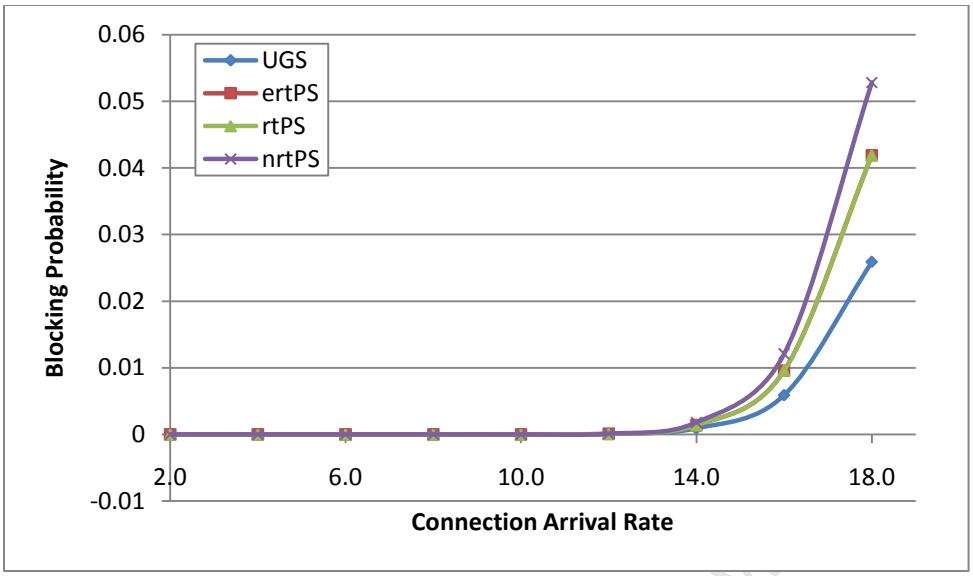

Figure 5.7: Blocking Probability of Connection types under Minimum bbu of nrtPS Connection

To fully understand the behaviour of each service type under different basic bandwidth requirements of nrtPS connection, Figure 5.8 to Figure 5.11 are presented. Figure 5.8 shows the behaviour of UGS connection with different bandwidth unit offered to nrtPS connection. It is shown that the lowest connection blocking probability for UGS connection is achieved when the nrtPS connection request is offered the minimum bandwidth; this is denoted as UGS-Min. The blocking probability at 18th arrival rate increases from 0.03 to 0.12 and 0.22 for UGS-Min to UGS-Avg and UGS-Max respectively.

Figure 5.9 shows the behaviour of ertPS connection under different nrtPS offered bandwidth unit. With minimum nrtPS offered bandwidth, the ertPS-Min connection maintains zero blocking probability until the 14th arrival rate while ertPS-Avg maintains zero blocking until the 12th arrival rate. For the ertPS-Max zero blocking probability is maintained until the 10th arrival rate. This implies that more connection requests of ertPS are admitted when only minimum bandwidth unit is offered to nrtPS connections and this leads to efficient resource utilization.

More connections of nrtPS service are admitted when a new nrtPS connection request is offered the minimum bandwidth requirement. In Figure 5.10, the blocking probability of nrtPS-Min at the 18th arrival rate is 0.05 while those of ertPS, Figure 5.9 and rtPS, Figure 5.11 are 0.04 and 0.04 respectively.

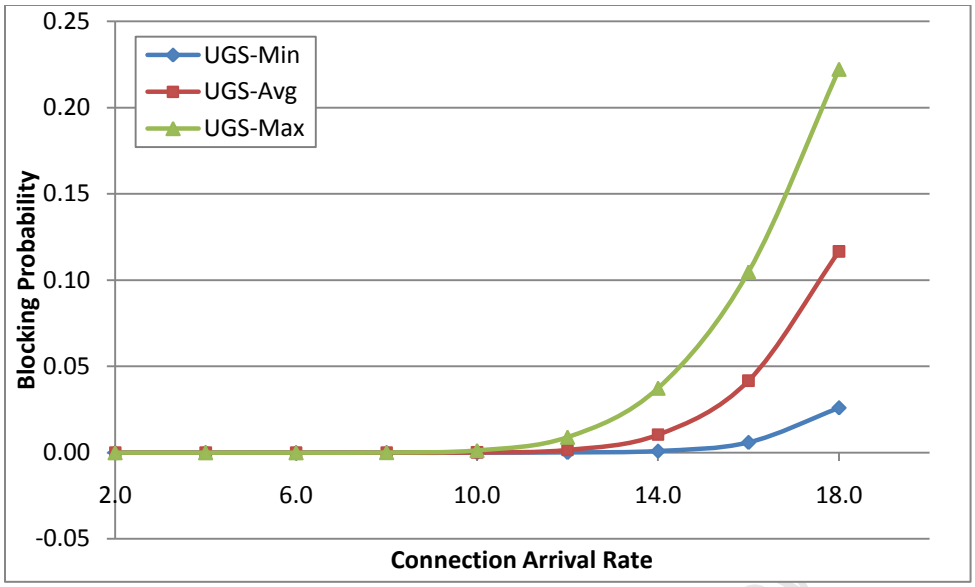

Figure 5.8: Blocking Probability of UGS Connections under different nrtPS bbus

It is seen from the presented results that when the bandwidth occupied by the nrtPS connections is degraded to the minimum bandwidth requirement, the degraded bandwidth is used to accept connection requests of all the service types, thereby more connection requests are admitted and connection blocking probability is reduced.

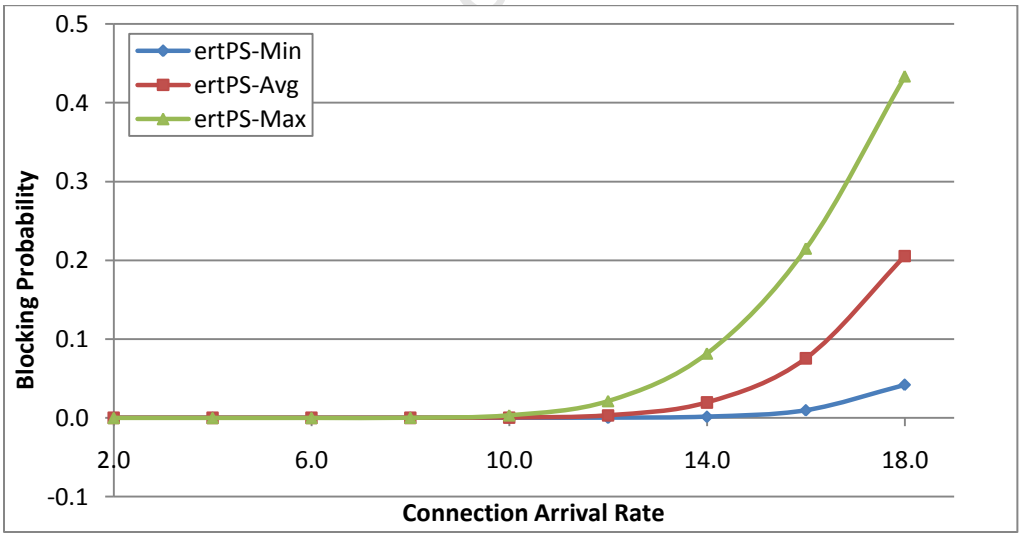

Figure 5.9: Blocking Probability of ertPS connections under different nrtPS bbus

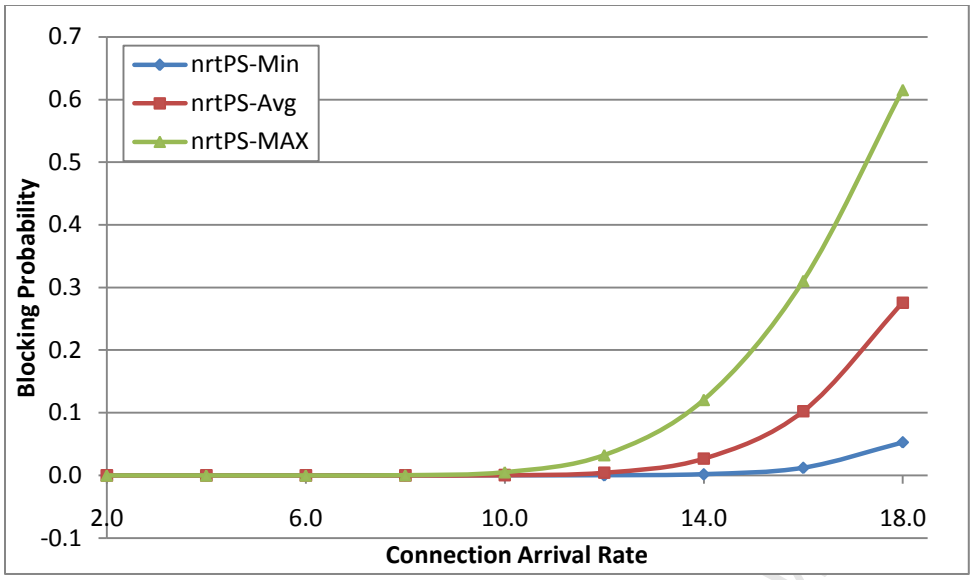

Figure 5.10: Blocking Probability of nrtPS Connections under different nrtPS bbus

5.4.1.3 Scenario 3

In this section we examine the connection throughput, the effective connection arrival rate of the connection requests into the system. Since we know that not all connection requests arrival are admitted, this section gives the picture of the behavior of each connection type as usual under different offered bandwidth unit of nrtPS connection.

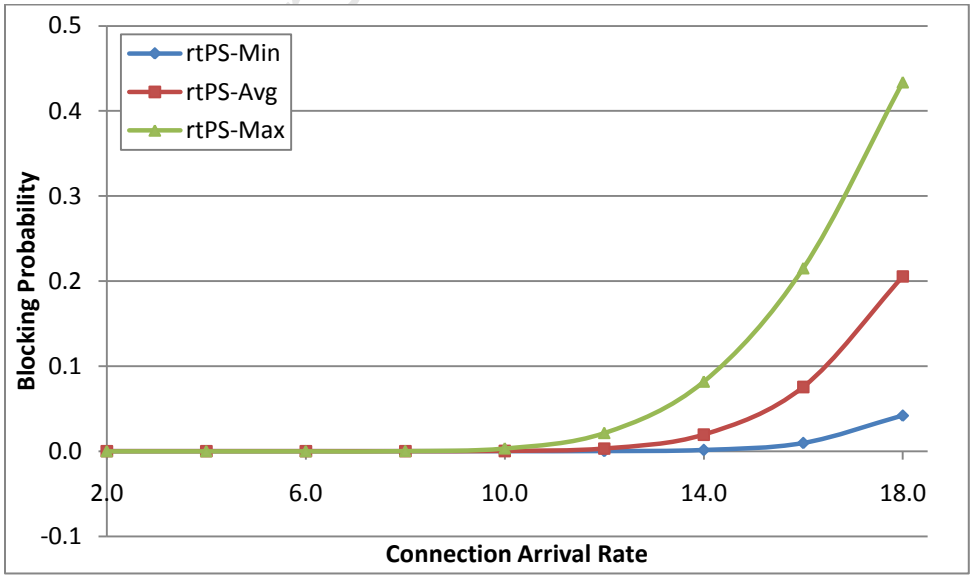

Figure 5.11: Blocking Probability of rtPS Connections under different nrtPS bbus

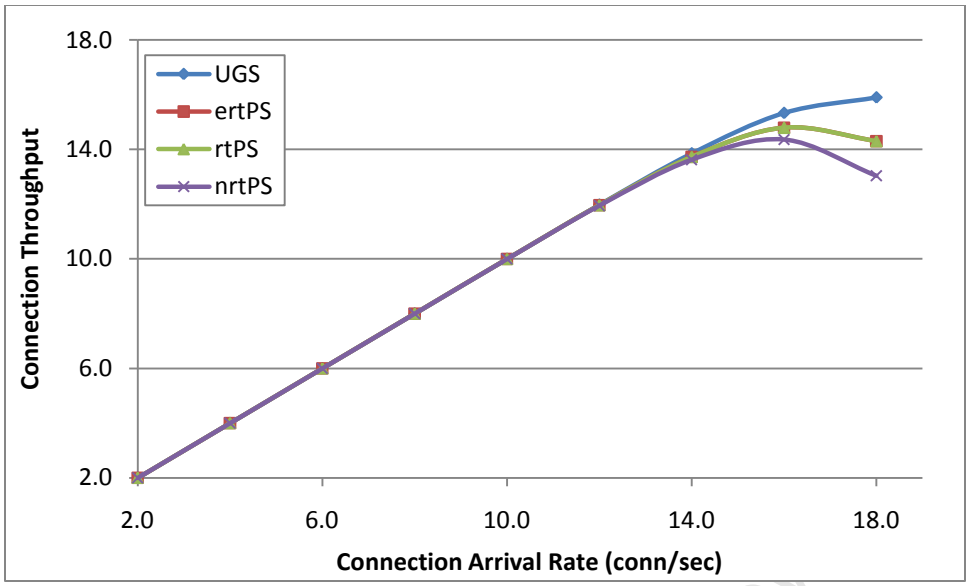

Figure 5.12: Connection Throughput vs. Connection Arrival Rate of Connection types

Figure 5.12 shows the connection throughput of UGS, ertPS, rtPS and nrtPS connections. It is shown that the connection throughput increase from zero to 16th arrival rate. Beyond this point more UGS connection requests can still be admitted due to the given threshold which is equivalent to the capacity of the system. However, throughput of ertPS, rtPS and nrtPS starts decreasing. The nrtPS connection suffers the highest drop in throughput.

The connection throughputs of the service types under different network loads of nrtPS service are presented in Figure 5.13 to Figure 5.15. In Figure 5.13, there is no drop in throughput achieve by UGS-Min which implies that more connection requests can still be admitted with guaranteed throughput. Likewise, UGS-Avg still achieves throughput increase meaning more connection requests can still be admitted with but not with guaranteed throughput. Beyond 16th arrival rate, the throughput of UGS-Max begins to decrease.

The connection throughput of rtPS-Min in Figure 5.14 shows an increase with connection arrival rate. This shows that more connection requests can still be admitted into the system. But with rtPS-Avg, throughput starts to drop after 16th arrival rate. The rtPS-Max experiences the highest drop in throughput because there is reduction in available resources to support the connection requests and the requests are blocked.

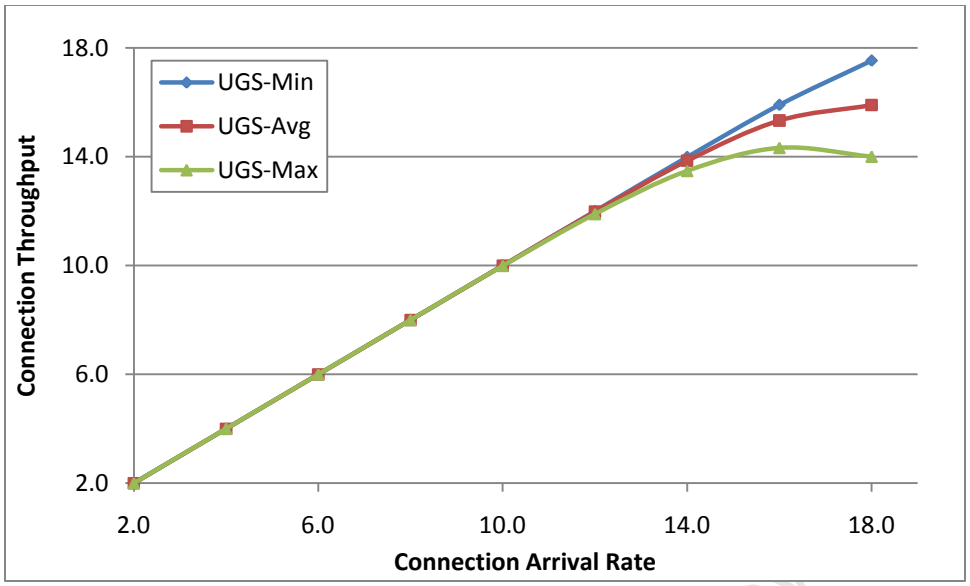

Figure 5.13: Connection Throughput of UGS Connections under different nrtPS bbus

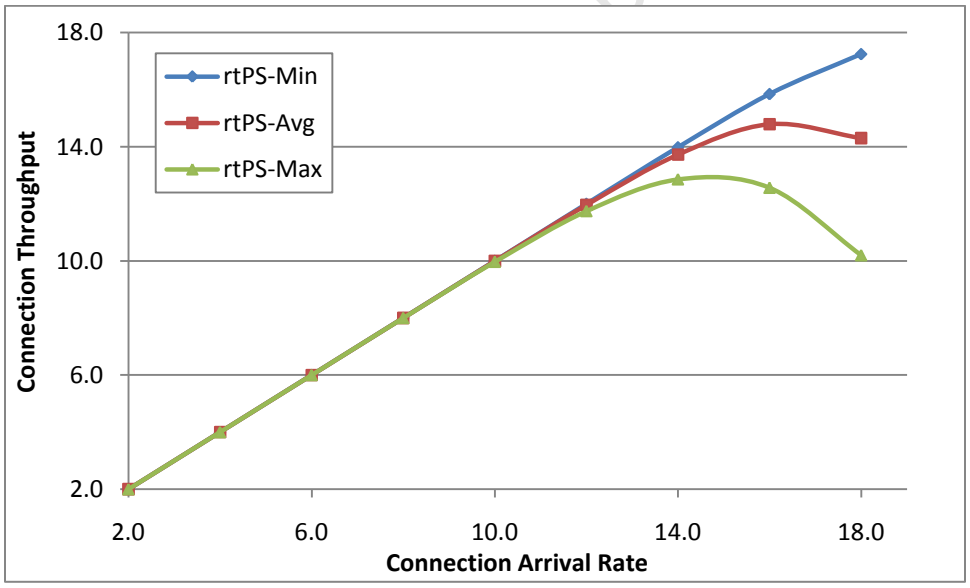

Figure 5.14: Connection Throughput of rtPS Connection under Different nrtPS bbus

In Figure 5.15, the throughput of nrtPS connection is presented. While connection throughput for nrtPS-Min increases with connection arrival rate which implies more connection requests can still be admitted with good throughput, the throughputs of nrtPS-Avg and nrtPS-Max start to decrease after some point. Beyond the 16th arrival rate throughput of connection requests start to decrease and this leads to an increase in blocking probability of connection

requests. For the maximum bandwidth unit offered to nrtPS, nrtPS-Max suffers the highest throughput drop hence highest dropping probability.

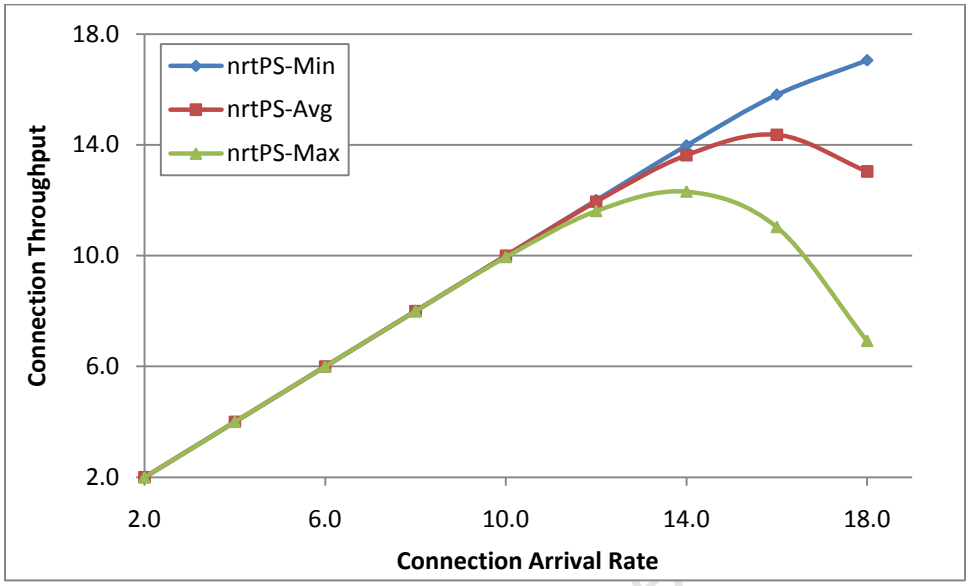

Figure 5.15: Connection Throughput of nrtPS Connections under different nrtPS bbus

The throughput of a new nrtPS connection request starts to drop after 14th arrival rate and drops to 7.0 conn/sec at 18th arrival rate.

5.4.2 Packet Scheduling

In this section we present the performance results of the proposed scheduling mechanism. A number of researches on traffic modeling have been carried out to investigate the characteristics of different traffic sources for various communication networks [43], [44] and [45]. Table 5.2 shows the arrival process and message size distribution of bandwidth requests of ertPS, rtPS, nrtPS and BE service types.

The bandwidth requests of ertPS are modeled as ON/OFF source model. When a source is ON, it generates packets with a constant inter-arrival time. When a source is off it does not generate any packets. The parameters of other service types are given in the Table 5.2. More explanation on Pareto distribution is given in Appendix B

When a bandwidth request message arrives to the base station, the request is queued in the appropriate local queue belonging to the connection service type. The request waits until the

higher priority service types have been scheduled before getting its service turn.

Table 5.2: Arrival Process and Message Size Distribution of the Traffic Sources with Priority Arrangement.

Traffic	Arrival Process	Message size distribution	Priority level
ertPS	Exponential ON/OFF: mean OFF period:1.67; mean ON period: 1.345	Deterministic, The size is 66 bytes	1
rtPS	Poisson, Mean interarrival time: 5s	Pareto cut-off: α=1.1, Minimum message = 4.5kbytes, Maximum message = 2Mbytes	2
nrtPS	Poisson, Mean interarrival time: 5s	Pareto cut-off: α=1.1, Minimum message = 4.5kbytes, Maximum message = 2Mbytes	3
BE	Poisson, Mean interarrival time: 7.5s	Exponential, mean=1900 bytes	4

The total delay suffered by the bandwidth request message is given by Equation (4.16). From the equation, we can calculate the mean message delay of rtPS, nrtPS and BE service types. Since the bandwidth request message of service type with highest priority is first served in all cases, the ertPS achieves the lowest mean message delay. The mean message delay of rtPS, nrtPS and BE are presented in Figure 5.16. It is seen from the Figure that both nrtPS and rtPS mean delays are lower compared to BE service which is delay tolerant.

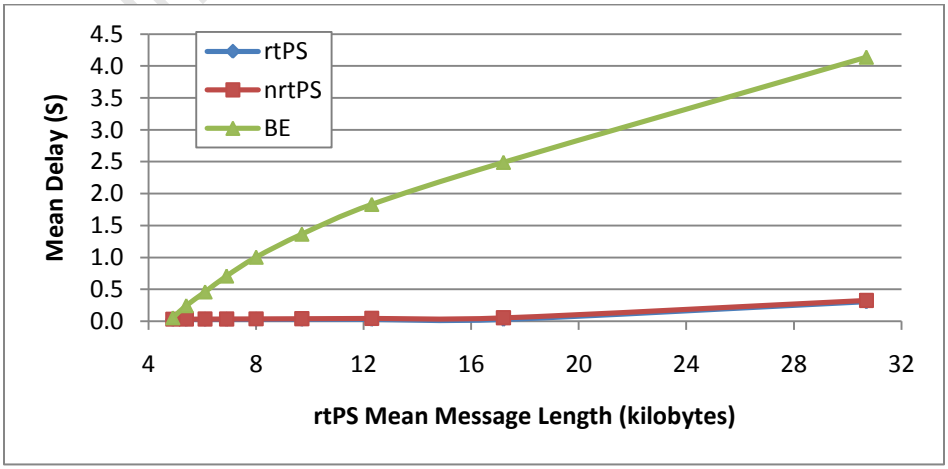

Figure 5.16: Mean Message delay of rtPS, nrtPS and BE Service Types

For clarity of comparison, Figure 5.17 shows the delay suffered by rtPS and nrtPS bandwidth requests. It is seen that rtPS suffered the lower delay.

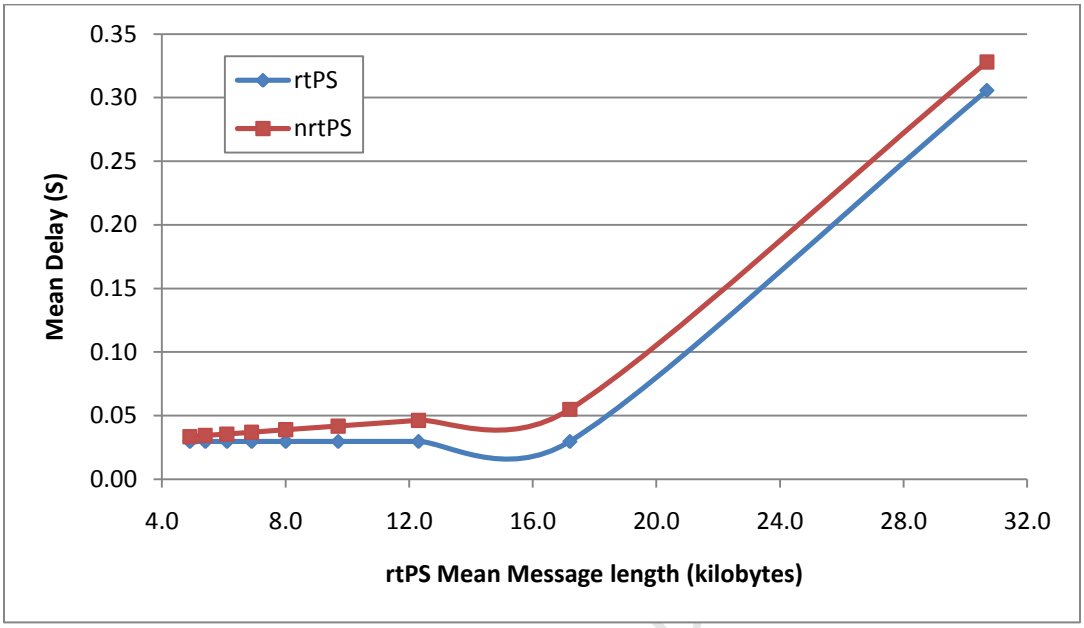

Figure 5.17: Mean Message delay of rtPS and nrtPS Service Types

Figure 5.18 shows the delay behavior of rtPS bandwidth request message with cut-off of $\alpha = 1.1$ for rtPS-A and $\alpha = 1.4$ for rtPS-B. Likewise, Figure 5.19 and Figure 5.20 show the effect of cut-off values of rtPS on nrtPS and BE respectively. The mean message delay decreases with increase in rtPS cut-off values.

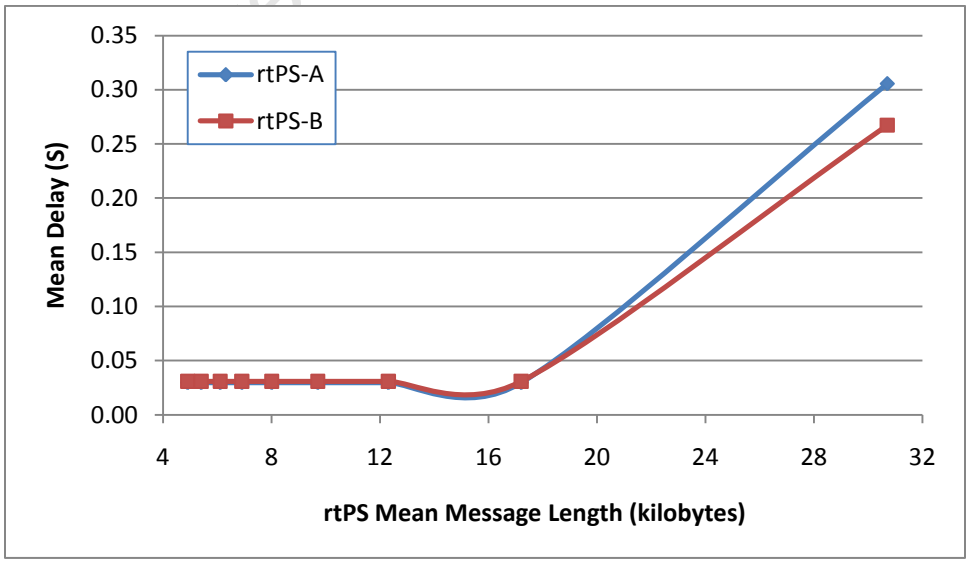

Figure 5.18: Mean Message delay of rtPS with different rtPS Cut-offs

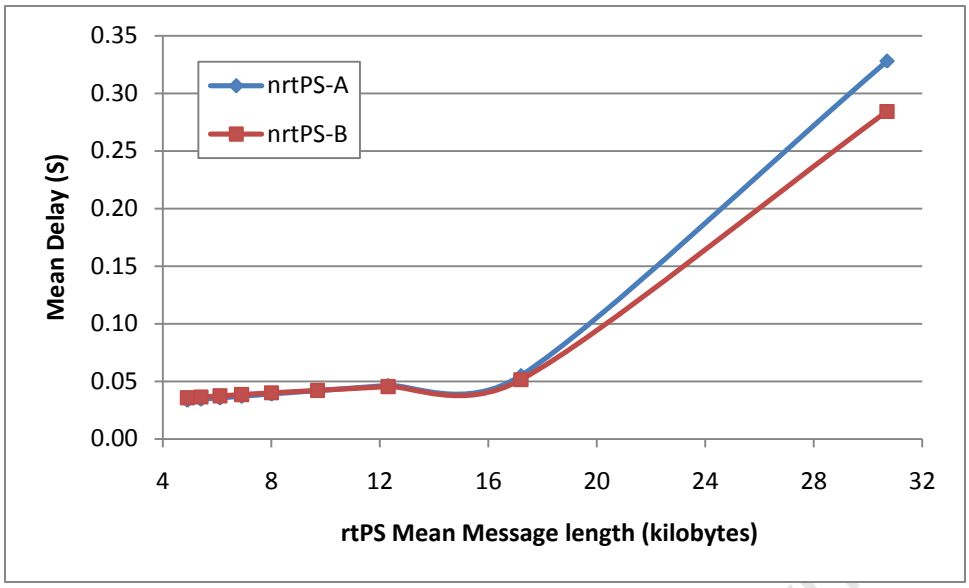

Figure 5.19: Mean Message Delay of nrtPS with different rtPS Cut-off

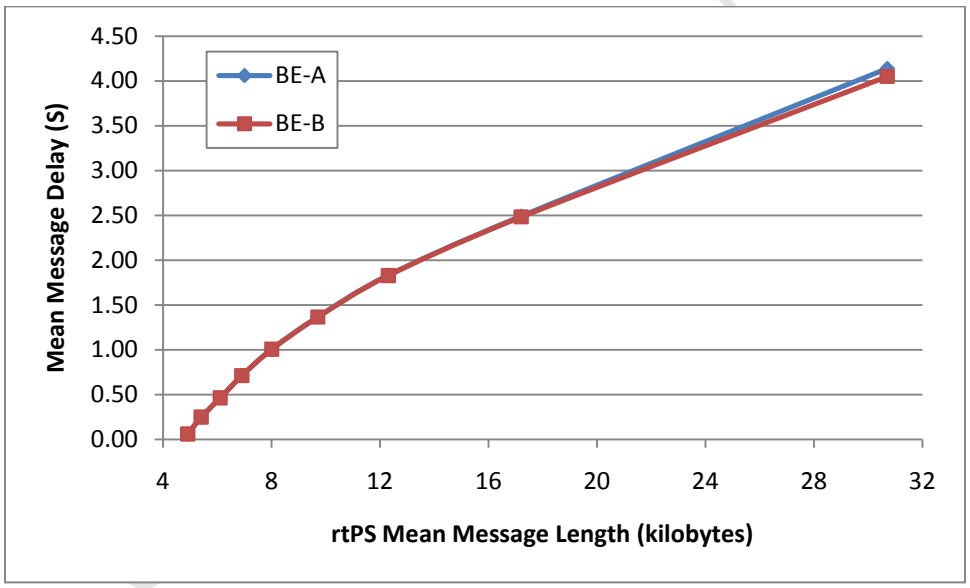

Figure 5.20: Mean Message Delay of BE with different rtPS Cut-offs

5.5 Chapter Discussion

This chapter has presented the performance evaluation of the framework developed in the previous chapters. Three scenarios were used to evaluate the performance of connection admission control while the performance of packet scheduler was evaluated using message delay.

In the first scenario, three CAC schemes were considered; the proposed Quadra-threshold

CAC scheme (QT), scheme without CAC (NC) and partitioning scheme (PS). The NC and PS schemes have been widely used in the literature. The blocking probabilities of UGS, ertPS, rtPS and nrtPS were considered under the three schemes. It was found from the results obtained that the proposed scheme achieved the lowest blocking probability when compared to the other schemes. This implied that admitted connection request would be offered request QoS.

The second scenario considered the effects of degrading the bandwidth of nrtPS ongoing connections in the network in other to admit more connection requests of the UGS, ertPS, rtPS and nrtPS service types. The bandwidth offered the on-going nrtPS connection was degraded from maximum bandwidth requirement to average bandwidth requirement and finally to minimum bandwidth requirement. The effect of bandwidth degradation of the on-going nrtPS connection was considered on admission of new connection requests of service type. The results obtained showed that more connection can be admitted into the system by degrading the offered bandwidth of ongoing nrtPS connections to their minimum required bandwidth. Thus, connection blocking probability was reduced.

The third scenario presented the actual throughput achieved by each service type under the degradation mechanism of nrtPS. Results showed that maximum throughput was achieved by each service type when the bandwidth offered on-going nrtPS connection was degraded to minimum bandwidth requirement.

Scheduling scheme was evaluated by using mean message delay metric. From the results obtained it was discovered that the scheduling scheme was able to guarantee the delay requirements of ertPS and rtPS connections.

Based on the results obtained from the performance evaluation in this chapter, conclusions and recommendations are presented in the next chapter.

Chapter 6 Conclusions and Recommendation

In this chapter, the summary of the project is given along with significant conclusions. Future work is also identified that could result from work done in this thesis.

In this thesis, the issues of admission control and packet scheduling for service types in IEEE 802.16 networks were addressed. The key contributions of this research were the development of admission control and packet scheduling algorithms for service differentiation and QoS support in IEEE 802.16 networks. A Quadra-threshold bandwidth-based connection admission control was proposed.

6.1 Summary

The Connection admission control algorithm made use of Quadra-threshold bandwidth sharing, bandwidth degradation and service class priority functionalities to make admission decisions. An analytical model based on the Markov decision process was developed for connection admission. Various service types of IEEE 802.16 networks were considered. A priority-based packet scheduling scheme adopting round robin multiprocessor sharing mechanism was proposed. Performance evaluation of the proposed schemes was carried out through MATLAB simulation. Performance metrics such as connection blocking probability and connection throughput were considered for connection admission control and delay metrics were considered for the scheduling algorithm. The proposed Quadra-threshold connection admission control scheme was compared with a generic bandwidth partitioning scheme and the scheme without connection admission control in terms of blocking probability and throughput. In addition, the performance of nrtPS, rtPS, ertPS and UGS service types were considered under degradation mechanism in terms of blocking probability.

6.2 Conclusions

Based on the findings the in preceding chapter, the following conclusions have been drawn:

- With Quadra-threshold bandwidth sharing, the IEEE 802.16 service types can be differentiated by allocating a separate bandwidth threshold to each service type. This type of differentiation according to service requirements and associated priority is very efficient. In

addition, customers can be differentiated by offering them different levels of QoS requirements

- The Quadra-threshold bandwidth based connection admission control scheme is suitable for changing pattern of number of connection requests of a service type arriving into the network. Different threshold levels can be assigned to each service type according to QoS requirement and service demand. This is more efficient than the scheme without admission control where the network is overloaded and the bandwidth partitioning scheme where a service type cannot access resources beyond the allocated partitioned even though other partitions are unused. The presented result showed that our proposed scheme achieved the lowest blocking probability when compared to bandwidth partitioning and the scheme without connection admission control.

- With bandwidth degradation, connection blocking probability can be minimized and more connection requests can be admitted into the network. When the offered bandwidth of ongoing nrtPS connections is reduced to the minimum bandwidth requirement, more bandwidth is released into the network. Consequently, more connection requests are admitted. The effect of bandwidth degradation of nrtPS connections from maximum bandwidth requirement to minimum bandwidth requirement on the other service types showed from the presented result that this mechanism achieved reduction in connection blocking probabilities. This was also demonstrated by connection throughput.

- It is important to prioritize service types with strict delay requirements so that these service types would guarantee the requested QoS.

6.3 Recommendations and future work

In this thesis, connection admission control and packet scheduling have been designed for the uplink transmission; however the scheme can be extended to downlink transmission. Moreover, although we evaluate the performance of the proposed schemes in IEEE 802.16 networks, the developed scheme can be used for OFMD/TDMA-based networks with various QoS requirements for different applications. The scheme is also suitable for future wireless networks, including cellular networks, IEEE 802.11 and IEEE 802.15 wireless networks.

In the proposed work, we have considered static users who are only mobile within the antenna sector of their subscriber station. Future work could consider mobile users that are not

limited to their correspondence subscriber stations. In this case, during connection admission algorithm design, consideration could be given to handover connections moving from one subscriber station to another. In addition, the effect of physical channel condition on the admission policy was not considered in the thesis and this could be an interesting area of future research which can be extended to be a cross-layer resource optimization.

The connection admission control and packet scheduling schemes have been considered separately for efficient QoS provisioning. It could be an interesting topic if future research would consider the integration of connection admission control and packet scheduling into a single mechanism.

References

[1] Radha, Krishna, Rao, G.S.V. and R. Radhamani. *WiMAX: A Wireless Technology Revolution* 2008.

[2] Anonymous "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," *IEEE Std 802. 16-2004 (Revision of IEEE Std 802. 16-2001)*, pp. 0_1-857, 2004.

[3] Anonymous "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1," *IEEE Std 802. 16e-2005 and IEEE Std 802. 16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802. 16-2004)*, pp. 0_1-822, 2006.

[4] Anonymous "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems," *IEEE Std 802. 16-2009 (Revision of IEEE Std 802. 16-2004)*, pp. C1-2004, 2009.

[5] S. Dhawan, "Analogy of promising wireless technologies on different frequencies: Bluetooth, WiFi, and WiMAX," in *Wireless Broadband and Ultra Wideband Communications, 2007. AusWireless 2007. the 2nd International Conference on*, 2007, pp. 14-14.

[6] A. Esmailpour and N. Nasser, "Packet scheduling scheme with quality of service support for mobile WiMAX networks," in *Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on*, 2009, pp. 1040-1045.

[7] J. Freitag, N. L. S. da Fonseca and J. F. de Rezende, "Admission control in IEEE 802.11 networks," in *Global Telecommunications Conference Workshops, 2004. GlobeCom Workshops 2004. IEEE*, 2004, pp. 258-265.

[8] WiMAX Forum. WiMAX™ technology forecast 2007Available: http://www.wimaxforum.org/technology/downloads/wimax_forum_wimax_forecasts_6_1_08.pdf.

[9] R. Attar, D. Ghosh, C. Lott, Mingxi Fan, P. Black, R. Rezaiifar and P. Agashe, "Evolution of cdma2000 cellular networks: multicarrier EV-DO," *Communications Magazine, IEEE*, vol. 44, pp. 46-53, 2006.

[10] Anonymous "IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements. - Part 15.1: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks (WPANs)," *IEEE Std 802. 15. 1-2005 (Revision of IEEE Std 802. 15. 1-2002)*, pp. 0_1-580, 2005.

[11] Anonymous "IEEE Standard for Information Technology-Telecommunications and

Information Exchange Between Systems-Local and Metropolitan Area Networks-Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Std 802. 11-2007 (Revision of IEEE Std 802. 11-1999)*, pp. C1-1184, 2007.

[12] C. Szabo and K. Farkas, "Planning wireless municipal networks based on wi-Fi/WiMAX mesh networks-applications, technologies and business models," in *IEEE International Conference on Communications*, Cape Town, South Africa, 23-27 May, 2010, .

[13] V. H. Muntean and M. Otesteanu, "WiMAX versus LTE - an overview of technical aspects for next generation networks technologies," in *Electronics and Telecommunications (ISETC), 2010 9th International Symposium on*, 2010, pp. 225-228.

[14] A. Esmailpour and N. Nasser, "A novel scheme for packet scheduling and bandwidth allocation in WiMAX networks," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1-5.

[15] A. Belghith and L. Nuaymi, "Comparison of WiMAX scheduling algorithms and proposals for the rtPS QoS class," in *Wireless Conference, 2008. EW 2008. 14th European*, 2008, pp. 1-6.

[16] A. Antonopoulos, C. Skianis and C. Verikoukis, "Traffic-aware connection admission control scheme for broadband mobile systems," in *GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference*, 2010, pp. 1-5.

[17] S. Chandra and A. Sahoo, "An efficient call admission control for IEEE 802.16 networks," in *Local & Metropolitan Area Networks, 2007. LANMAN 2007. 15th IEEE Workshop on*, 2007, pp. 188-193.

[18] C. Prasun and S. M. Lti. A fair and efficient packet scheduling scheme for IEEE 802.16 broadband wireless access systems. *International Journal of Ad Hoc, Sensor &Ubiquitous Computing (IJASUC) 1(3)*, pp. 93-104. 2010.

[19] Chakchai So-In, R. Jain and A. -. Tamimi, "Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey," *Selected Areas in Communications, IEEE Journal on*, vol. 27, pp. 156-171, 2009.

[20] H. A. Chan, "From current to future wireless networks," in *Portable Information Devices, 2008 and the 2008 7th IEEE Conference on Polymers and Adhesives in Microelectronics and Photonics. PORTABLE-POLYTRONIC 2008. 2nd IEEE International Interdisciplinary Conference on*, 2008, pp. 1-6.

[21] J. Veijalainen and W. Rehmat, "Mobile communities in developing countries," in *Mobile Data Management (MDM), 2010 Eleventh International Conference on*, 2010, pp. 314-319.

[22] Anonymous "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and

Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1," *IEEE Std 802. 16e-2005 and IEEE Std 802. 16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802. 16-2004)*, pp. 0_1-822, 2006.

[23] A. Flizikowski, R. Kozik, M. Majewski and M. Przybyszewski, "Evaluation of guard channel admission control schemes for IEEE 802.16 with integrated nb-LDPC codes," in *Ultra Modern Telecommunications & Workshops, 2009. ICUMT '09. International Conference on*, 2009, pp. 1-8.

[24] H. Wang, W. Li and D. P. Agrawal, "Dynamic admission control and QoS for 802.16 wireless MAN," in *Wireless Telecommunications Symposium, 2005*, 2005, pp. 60-66.

[25] S. Kalikivayi, I. S. Misra and K. Saha, "Bandwidth and delay guaranteed call admission control scheme for QOS provisioning in IEEE 802.16e mobile WiMAX," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 2008, pp. 1-6.

[26] Liping Wang, Fuqiang Liu, Yusheng Ji and Nararat Ruangchaijatupon, "Admission control for non-preprovisioned service flow in wireless metropolitan area networks," in *Universal Multiservice Networks, 2007. ECUMN '07. Fourth European Conference on*, 2007, pp. 243-249.

[27] Lang Xie, Jie Xiang, Yan Zhang and Jin Zhang, "Joint bandwidth reservation and admission control in IEEE 802.16e based networks," in *Communications, 2009. ICC '09. IEEE International Conference on*, 2009, pp. 1-6.

[28] I. C. Msadaa, D. Câmara and F. Filali, "Scheduling and CAC in IEEE 802.16 Fixed BWNs: A Comprehensive Survey and Taxonomy," *Communications Surveys & Tutorials, IEEE*, vol. 12, pp. 459-487, 2010.

[29] D. Ndiki, H. J. Helgert and S. Hussein, "A comparative overview IEEE 802.16e QoS scheduling algorithms," in *Evolving Internet (INTERNET), 2010 Second International Conference on*, 2010, pp. 74-79.

[30] S. Alexanda, A. Olli, K. Juha and H. Timo. Ensuring the QoS requirements in 802.16 scheduling. Presented at MSWiM '06 Proceedings of the 9th ACM International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems. 2006, .

[31] E. Laias, I. Awan and P. M. L. Chan, "Fair and latency aware uplink scheduler in IEEE 802.16 using customized deficit round robin," in *Advanced Information Networking and Applications Workshops, 2009. WAINA '09. International Conference on*, 2009, pp. 425-432.

[32] Jianfeng Chen, Wenhua Jiao and Hongxi Wang, "A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode," in *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, 2005, pp. 3422-3426 Vol. 5.

[33] K. Wongthavarawat and A. Ganz, "Packet Scheduling for QoS Support in IEEE 802.16Broadband Wireless Access Systems," *International Journal on Communication Systems,*

vol. 16, pp. 81-96, 2003.

[34] J. F. Borin and N. da Fonseca, "Uplink scheduler and admission control for the IEEE 802.16 standard," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE,* 2009, pp. 1-6.

[35] D. S. Shu'aibu and Y. Syed S.K., "Link Aware Call Admission and Packet Scheduling for Best Effort and UGS Traffics in mobile WiMAX," *International Journal of the Physical Sciences.,* vol. 6(7), pp. 1694-1701, 4 April, 2011.

[36] Chi-Hong Jiang and Tzu-Chieh Tsai, "Token bucket based CAC and packet scheduling for IEEE 802.16 broadband wireless access networks," in *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE,* 2006, pp. 183-187.

[37] P. G. Potter and M. Zukerman, "Analysis of a discrete multipriority queueing system involving a central shared processor serving many local queues," *Selected Areas in Communications, IEEE Journal on,* vol. 9, pp. 194-202, 1991.

[38] J. S. William, *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modelling.* 2009.

[39] Ke Yu, Xuan Wang, Songlin Sun, Lin Zhang and Xiaofei Wu, "A statistical connection admission control mechanism for multiservice IEEE 802.16 network," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th,* 2009, pp. 1-5.

[40] S. Kalikivayi, I. S. Misra and K. Saha, "Bandwidth and delay guaranteed call admission control scheme for QOS provisioning in IEEE 802.16e mobile WiMAX," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE,* 2008, pp. 1-6.

[41] The Math Works, "MATLAB User's Guide ," 2009.

[42] S. K. Falowo and N. Ventura, "Connection admission control (CAC) for QoS differentiation in PMP IEEE 802.16 networks," in *AFRICON, 2011,* 2011, pp. 1-6.

[43] P. Latkoski and L. Gavrilovska, "Web traffic over bluetooth: Modeling, analysis and performance evaluation," in *Wireless and Mobile Communications, 2007. ICWMC '07. Third International Conference on,* 2007, pp. 71-71.

[44] Z. Sun, D. He, L. Liang and H. Cruickshank, "Internet QoS and traffic modelling," *Software, IEE Proceedings -,* vol. 151, pp. 248-255, 2004.

[45] R. Y. Wang, M. Zukerman and R. J. Harris, "Modelling PTT packet delay in the GPRS/GSM uplink," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd,* 2006, pp. 420-424.

Appendix A: IEEE 802.16 QoS enhancements

In this other IEEE 802.16 QoS enhancements are explained. These QoS enhancements include connection and service flows, data unit and bandwidth request and allocation.

Connection and Service flows

The MAC layer is connection oriented and identifies a logical connection between a BS and a SS by a 16-bit unidirectional connection identifier (Connection ID – CID). The BS uses the CID to identify the connection between the peer MAC/PHY entities and for carrying data and control plane traffic. The CIDs for uplink and downlink connections are different. There is a unique CID for a given BS/SS pairing and changes when a SS moves from one BB to another. There are two types of connections: transport and management connection. The transport connection is used for data transmission while the management connection is used for management related functions.

A service flow is a 32-bit number called service flow identifier (SFID) that defines a connection through a set of QoS parameters. A SFID is mapped to a unique CID and the base station maintains the association between the two identifiers. A SS can have multiple service flows. Service flow between both UL and DL direction may exist without actually being activated to carry traffic; however, a CID corresponds to an active flow.

Data Unit

Two types of data unit are defined in MAC layer: protocol data unit (PDU) and service data unit (SDU).

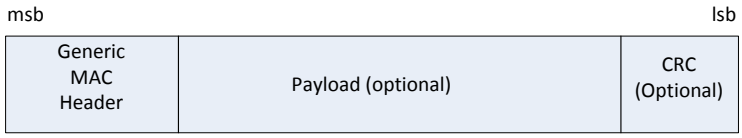

Figure 0.1: MAC PDU

The MAC PDU (Figure 0.1) is the data unit exchanged between the MAC layers of a BS and a SS. PDUs are exchanged among peer entities in the same protocol layer from higher to lower layers in the downward direction at the sender side and from lower to higher layers in the

upward direction at the receiver side (Figure 0.3). Each PDU begins with a MAC header, an optional payload and a cyclic redundancy check (CRC).

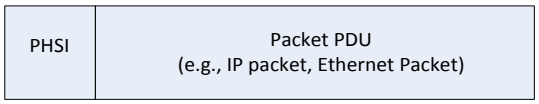

Figure 0.2: MAC SDU

The MAC SDU (Figure 0.2) is the data unit exchange between two adjacent protocol layers. On the downward direction, it is the data unit received from the previous higher layer. On the upward direction, it is the data unit sent to the next higher layer (Figure 0.3).

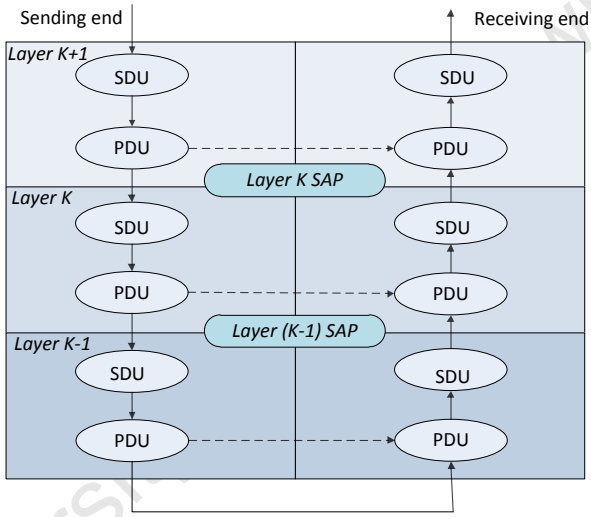

Figure 0.3: PDU and SDU in a Protocol Stack

To maximize the efficiency of transmission, MAC PDUs are constructed using concatenation, fragmentation and packing. In concatenation, multiple MAC PDUs are combined to form one PDU for transmission. The process in which a MAC SDU is divided into one or more MAC PDUs is referred to as fragmentation. Packing involves the packing of multiple MAC SDUs into a single MAC PDU. Two types of MAC PDU are defined: generic MAC PDU and bandwidth request PDU. Generic MAC PDU is for carrying data and MAC layer signalling messages. It starts with a generic header and follows with payload and optional cyclic redundancy check (CRC). Subscriber stations use bandwidth request PDUs to inform BS of their bandwidth requirements in the UL, due to pending data transmission. A bandwidth request PDU consists of a bandwidth request header without a payload or CRC.

Bandwidth Request and Allocation

In the downlink, all decisions related to the allocation of bandwidth to various SSs are made by the BS on a per connection basis, which does not require the involvement of the SSs. As MAC data units arrive for each connection, the BS schedules them for the PHY resources, based on their QoS requirements. Once dedicated PHY resources have been allocated for the transmission of the PDU, the BS indicates this allocation to the SS, using DL-MAP message.

In the uplink, the SS makes bandwidth request which is piggybacked onto a generic MAC PDU. Since the burst profile associated with a connection can change dynamically, all bandwidth requests are made in terms of bytes of information. The message specifies the specific connection requesting for bandwidth and the amount of bandwidth requested. Bandwidth request in the uplink can be incremental or aggregate requests.

In general, uplink bandwidth request is made per connection while bandwidth grant is performed in two ways: grant per connection (GPC) and grant per subscriber station (GPSS). In GPC, BS MAC scheduler handles each connection request of the SSs independently and the bandwidth is explicitly granted to each connection while in GPPS, when multiple bandwidth requests are associated with a particular SS, aggregate bandwidth is granted to the SS station. A scheduler needs to be implemented within the SS MAC to allocate the granted bandwidth to all or some of its connections based on the amount of pending traffic and their QoS requirements.

Appendix B: Pareto Distribution

One of the heavy-tailed distributions is the Pareto distribution, which is the power law over the entire range. The Pareto distribution is useful where occurrences of small values are common and occurrences of large value are rare.

The following formulas describe the most important functions of the distribution

Cumulative Distribution Function (CDF):

$$y = F(x) = 1 - \left(\frac{k}{x}\right)^{\alpha} \quad \textbf{(0.1)}$$

The parameter k is the minimum value of x.

The mean of the normal Pareto distribution without cut-off is given by:

$$\mu = \frac{k\alpha}{\alpha - 1} \quad \alpha > 1$$

If the maximum value of x is limited to be m, the mean of the Pareto distribution with cut-off becomes

$$\mu_c = \frac{\alpha k - m\left(\frac{k}{m}\right)^{\alpha}}{\alpha - 1} \quad \alpha > 1$$

If uniformly distributed random variable y (between 0 and1) is used, x becomes a random variable that follows Pareto distribution and x is found as:

$$x = \frac{k}{e^{\frac{\log_e(1-y)}{\alpha}}}$$

Appendix C: Hardware and Software Specifications

Hardware

Operating System:	Windows XP Professional (5.1, Build 2600) Service Pack 3
System Model:	MS-7529
Model Name:	Pentium (R) Dual-Core CPU E5400 @2.70GHz (2 CPUs)
Memory:	2038MB RAM
Family:	6
Model:	23
Stepping:	10
Vendor_id:	GenuineIntel

Software

Name:	MATLAB
Version:	7.9.0.529 (R2009b), 32-bit (Win 32)

Appendix D: Accompany CD-ROM

The CD-ROM included with this thesis contains the following files and information:

- *Research Literature* – Electronic copies of the research papers and other literature used during the course of this research can be found in the directory labelled "Research Literature".
- *Software* – All the software code developed for the evaluation framework can be found in the directory labelled "Software".
- *Publications* – Copies of papers which have been accepted to conferences can be found in the directory "Publications".
- *Thesis* – An electronic copy, in PDF format, of this document can be found in the directory labelled "Thesis".
- *Results* – The results obtained during the performance tests carried out for the thesis can be found in the directory labelled "Results".